DATA PROCESSING METHOD AND

STORAGE MEDIUM, AND PROGRAM FOR CAUSING

COMPUTER TO EXECUTE THE DATA PROCESSING METHOD

BACKGROUND OF THE INVENTION

5

The present invention relates to a media content data processing method, a storage medium, and a program, all being related to viewing, playback, and delivery of continuous audio-visual data (media content), such as a motion picture,

10 a video program, or an audio program, wherein the synopsis or a highlight scene of media content or only a scene of media content desired by the audience is played back and delivered.

Conventional media content has conventionally been played back, delivered, or stored on the basis of individual files

15 storing media content.

As described in Japanese Patent Laid-Open No. Hei-10-111872, according to a method of retrieving a specific scene of a motion picture, a change between scenes of the motion picture (hereinafter referred to as a "scene cut") is detected. To each

20 scene cut are added additional data, such as a time code of the start frame, a time code of the end frame, and a keyword of the scene.

As an alternative method, Carnegie Mellon University (CMU) has attempted to summarize a motion picture by detecting scene

25 cuts of a motion picture, detecting a human face or a caption,

and detecting a key phrase through speech recognition [Michael A. Smith and Takeo KANADE "Video Skimming and Characterization through Combination of Image and Language Comprehension Techniques" CMU-CS-97-111, February 3, 1997].

5   When the motion picture is played back on a per-file basis, reviewing the synopsis of the motion picture has been impossible. Further, even when a highlight scene or scenes desired by the user are retrieved, the scene or scenes must be searched from the head of media content. Further, in the case of delivery

10   of a motion picture, all the data sets of a file are transmitted, thus requiring a very long transmission time.

According to the method described in Japanese Patent Application Laid-Open No. Hei-10-111872, scenes can be retrieved through use of a keyword, thus facilitating retrieval of scenes

15   desired by the user. The additional data do not include a relationship or connection between the scenes. For this reason, the method encounters difficulty in retrieving, e.g., one subplot of a story. Further, when retrieving scenes based on only a keyword, the user encounters difficulty in gaining awareness

20   of which scenes are contextually important. Therefore, preparation of a synopsis or highlight scenes becomes difficult.

The method developed by CMU enables summarization of a motion picture. However, summarization results in a digest of a single, fixed pattern. For this reason, summarization of a

25   motion picture into a digest which requires a different playback

2

time; for example, a digest whose playback time assumes a length of three or five minutes, is difficult. Further, summarization of a motion picture desired by the user; such as selection of scenes including a specific character, is also difficult.

₅ ## SUMMARY OF THE INVENTION

The object of the present invention is to provide means for selecting, playing back, and delivering only a synopsis, a highlight scene, or a scene desired by the audience, at the time of playback of media content.

10   Another object of the present invention is to provide means for playing back a synopsis, a highlight scene, or a scene desired by the audience within a period of time desired by the user, at the time of selection of the synopsis, the highlight scene, or the desired scene.

15   Still another object of the present invention is to provide means for delivering only a synopsis, a collection of high-light scenes, or a scene desired by the user, within a period of time desired by the user, at the request of the user during the delivery of media content.

20   Yet another object of the present invention is to provide means for controlling the amount of data to be delivered, in accordance with the traffic volume of a line through which the user establishes communication with a server.

To solve problems of the prior art, according to one aspect

25 of the present invention, there is provided a data processing

3

method comprising the steps of: inputting context description data described in a hierarchical structure, wherein the hierarchical structure comprises the highest hierarchical layer in which time-varying media content and the context of the media content are formed into a single element representing media content; the lowest hierarchical layer in which an element represents a media segment formed by dividing the media content and is assigned, as an attribute, time information relating to a corresponding media segment and a score; and other hierarchical layers include elements which are directly or indirectly associated with at least one of the media segments and which represent scenes or a set of scenes; and selecting at least one segment from the media content, on the basis of the score assigned to the context description data.

Preferably, the data processing method further comprises an extraction step for extracting only data corresponding to the segment selected by the selection step, from the media content.

Preferably, the data processing method further comprises a playback step for playing back only data corresponding to the segment selected by the selection step, from the media content.

Preferably, the sore represents a contextual importance of media content.

Preferably, the score represents the degree of contextual importance of a scene of interest from the viewpoint of a keyword,

4

and in the selection step there is selected a scene in which the score is used from at least one viewpoint.

Preferably, the media content corresponds video data or audio data.

5   Preferably, the media content corresponds to data comprising video data and audio data, which are mutually synchronized.

Preferably, the context description data describe the configuration of video data or audio data.

10   Preferably, the context description data describe the configuration of each of video data sets and audio data sets.

Preferably, in the selection step, a scene is selected by reference to context description data pertaining to video data or audio data.

15   Preferably, the selection step comprises a video selection step for selecting a scene of video data by reference to context description data of video data or an audio selection step for selecting a scene of audio data by reference to context description data of audio data.

20   Preferably, the selection step comprises a video the selection step for selecting a scene of video data by reference to context description data of video data, and an audio selection step for selecting a scene of audio data by reference to context description data of audio data.

25   Preferably, the data to be extracted in the extraction step

5

correspond to video data or audio data.

Preferably, the data to be extracted in the extraction step correspond to data comprising video data and audio data, which are mutually synchronized.

5   Preferably, media content comprises a plurality of different media data sets within a single period of time. Further, the data processing method comprises a determination step of receiving structure description data having a data configuration of the media content described therein and determining which

10   one of the media data sets is to be taken as an object of selection, on the basis of determination conditions to be used for determining data as an object of selection. Further, in the selection step, data are selected from only the data sets, which have been determined as objects of selection by the determination

15   means, by reference to the structure description data.

Preferably, the data processing method further comprises: a determination for receiving structure description data having a data configuration of the media content described therein and determines whether only video data, only audio data, or both

20   video data and audio data are taken as an object of selection, on the basis of determination conditions to be used for determining data as an object of selection. Further, in the selection step, data are selected from only the data sets determined as objects of selection by the determination step,

25   by reference to the structure description data.

6

Preferably, media content comprises a plurality of different media data sets within a single period of time. Preferably, in the determination step, there are received structure description data having a data configuration of the media content described therein, and a determination is made as to which one of the video data sets and/or audio data sets is to be taken as an object of selection. Further, in the selection step, data are selected from only the data sets determined as objects of selection by the determination step, by reference to the structure description data.

Preferably, representative data pertaining to a corresponding media segment are added, as an attribute, to individual elements of context description data in the lowest hierarchical layer; and in the selection step, there are selected the entire data pertaining to the media segment and/or representative data pertaining to a corresponding media segment.

Preferably, the entire data pertaining to the media segment correspond to media data, and the media content comprises a plurality of different media data sets within a single period of time. Preferably, the data processing method further comprises a determination step for receiving structure description data having a data configuration of the media content described therein and determining which one of the media data sets and/or representative data sets is to be taken as an object of selection. Further, in the selection step, data are selected

from only the data sets determined as objects of selection by the determination step, by reference to the structure description data.

Preferably, the data processing method further comprises:

5 a determination step for receiving structure description data having a data configuration of the media content described therein and determining whether only the entire data pertaining to the media segment, only the representative data pertaining to the media segment, or both the entire data and the

10 representative data pertaining to a corresponding media segment are to be taken as objects of selection, on the basis of determination conditions to be used for determining data as an object of selection. Further, in the selection step, data are selected from only the data sets determined as objects of

15 selection by the determination means, by reference to the structure description data.

Preferably, the determination conditions comprise at least one of the capability of a receiving terminal, the traffic volume of a delivery line, a user request, and a user's taste, or a

20 combination thereof.

Preferably, the data processing method further comprises a formation step for forming a stream of media content from the data extracted by the extraction step.

Preferably, the data processing method further comprises

25 a delivery step for delivering the stream formed by the formation

8

step over a line.

Preferably, the data processing method further comprises a recording step for recording the stream formed by the formation step on a data recording medium.

5   Preferably, the data processing method further comprises a data recording medium management step for re-organizing the media content that has already been stored and/or media content to be newly stored, according to the available disk space of the data recording medium.

10   Preferably, the data processing method further comprises a stored content management step for re-organizing the media content stored in the data recording medium according to the period of storage of the media content.

..According to yet another aspect of the present invention, 15   there is provided a computer-readable recording medium on which the previously-described data processing method is recorded in the form of a program to be performed by a computer.

According to still another aspect of the present invention, there is provided a program for causing a computer to perform 20   the previously-described data processing method.

In the data processing method, the recording medium, and the program of the present invention, selection means (corresponding to a selection step) selects at least one segment from a media content on the basis of a score appended, as an 25   attribute, to the lowest hierarchical layer or other hierarchical

9

layers of context description data, through use of context description data of hierarchical stratum which comprises the highest hierarchical layer, the lowest hierarchical layer, and other hierarchical layers obtained by input means (corresponding

5   to an input step).

Particularly, the extraction means (corresponding to the extraction step) extracts only the data pertaining to a segment selected by the selection means (corresponding to the selection step).

10 Particularly, the playback means (corresponding to the playback step) plays back only the data pertaining to the segment selected by the selection means (corresponding to the selection step) are played back.

Accordingly, a more important scene can be freely selected

15 from the media content, and the thus-selected important segment can be extracted or played back. Further, the context description data assume a hierarchical stratum comprising the highest hierarchical layer, the lowest hierarchical layer, and other hierarchical layers. Scenes can be selected in arbitrary

20 units, such as on a per-chapter basis or a per-section basis. There may be employed various selection formats, such as selection of a certain chapter and deletion of unnecessary paragraphs from the chapter.

In the data processing method, the recording medium, and

25 the program of the present invention, a score represents the

10

degree of contextual importance of media content. So long as the score is set so as to select important scenes, a collection of important scenes of a program, for example, can be readily prepared.

5   Further, so long as the score is set so as to represent the importance of a scene of interest from the viewpoint of keyword, segments can be selected with a high degree of freedom by determination of a keyword. For example, so long as a keyword is determined from a specific viewpoint, such as a character

10   or an event, only the scenes desired by the user can be selected.

In the data processing method, the recording medium, and the program of the present invention, the media content corresponds to video data and/or audio data, and the context description data describe the configuration of respective video

15   data sets and/or audio data sets. The video selection means (corresponding to the video the selection step) selects a scene by reference to the context description data pertaining to video data. The audio selection means (corresponding to the audio the selection step) selects a scene by reference to the context description data pertaining to audio data.

20   description data pertaining to audio data.

Further, the extraction means (corresponding to the extraction step) extracts video data and/or audio data.

An important segment can be selected from the video data and/or audio data, and video data and/or audio data pertaining

25   to the thus-selected segment can be extracted.

11

In the data processing method, the recording medium, and the program of the present invention, in a case where media content comprises a plurality of different media data sets within a single period of time, the determination means (corresponding to the determination step) determines which of the media data sets is to be taken as an object of selection, on the basis of determination conditions to be used for determining data as an object of selection. The selection means (corresponding to the selection step) selects data set from only the data determined by the determination means (corresponding to the determination step).

The determination conditions comprise at least one of the capability of a receiving terminal, the traffic volume of a delivery line, a user request, and a user's taste, or a combination thereof. For instance, the capability of a receiving terminal corresponds to video display capability, audio playback capability, or a rate at which compressed data are to be decompressed. The traffic volume of a delivery line corresponds to the degree of congestion of a line.

In a case where media content is divided into; for example, channels and layers and different media data sets are assigned to the channels and layers, the determination means (corresponding to the determination step) can determine media data pertaining to an optimum segment according to determination conditions. Accordingly, the selection means (corresponding to the selection step) can select an appropriate amount of media

12

data. In a case where channels and layers are employed as optimum segments, video data having a standard resolution may be assigned to a channel-1/layer-1 for transporting a motion picture, and video data having a high resolution may be assigned to a

5   channel-1/layer-2. Further, stereophonic data may be assigned to a channel-1 for transporting sound data, and monophonic data may be assigned to a channel-2.

In the data processing method, the recording medium, and the program of the present invention, the determination means

10   (corresponding to the determination step) determines whether only the video data, only the audio data, or both video and audio data are to be taken as an object of selection, on the basis of the determination conditions.

Before the selection means (corresponding to the selection

15   step) selects a segment, the determination means (corresponding to the determination step) determines which one of the media data sets is to be taken as an object of selection or whether only the video data, only the audio data, or both video and audio data are to be taken as an object of selection. As a result,

20   the time required by the selection means (corresponding to the selection step) for selecting a segment can be shortened.

In the data processing method, the recording medium, and the program of the present invention, representative data are appended, as an attribute, to individual elements of the context

25   description data in the lowest hierarchical layer, and the

13

selection means selects the entire data pertaining to a media segment and/or representative data pertaining to a corresponding media segment.

In the data processing method, the recording medium, and the program of the present invention, the entire data pertaining to a media segment correspond to media data, and the media content comprises a plurality of different media data sets within a single period of time. The determination means (corresponding to the determination step) determines which one of the media data sets and/or representative data are to be taken as objects of selection, on the basis of structure description data and determination conditions.

The media content is divided into; for example, channels and layers, and different media data sets are assigned to the channels and layers. The determination means can determine media data pertaining to an optimum segment (channel or layer) according to these determination conditions.

In the data processing method, the recording medium, and the program of the present invention, the determination means (corresponding to the determination step) determines whether only the entire data pertaining to a corresponding media segment, only the representative data pertaining to the corresponding media segment, or both the entire data and the representative data pertaining to the corresponding media segment are to be taken as objects of selection, on the basis of determination

14

conditions.

Before the selection means (corresponding to the selection step) selects a segment, the determination means (corresponding to the determination step) determines which one of the media data sets is to be taken as an object of selection or whether only the entire data or only the representative data, or both the entire data and the representative data are to be taken as objects of selection. As a result, the time required by the selection means (corresponding to the selection step) for selecting a segment can be shortened.

In the data processing method, the recording medium, and the program of the present invention, formation means (corresponding to the formation step) forms a stream of media content from the data extracted by the extraction means (corresponding to the extraction step). Accordingly, a stream or file which describes a piece of content corresponding to the thus-selected segment can be prepared.

In the data processing method, the recording medium, and the program of the present invention, the delivery means (corresponding to the delivery step) delivers the stream formed by the formation means (corresponding to the formation step) over a line. Therefore, data pertaining to only important segments can be delivered to the user.

In the data processing method, the recording medium, and the program of the present invention, the data recording medium

15

management means (corresponding to the data recording medium management step) re-organizes the media content that has been stored so far and/or media content to be newly stored, according to the available disk space of the data recording medium.

5   Particularly, in the data processing method, the recording medium, and the program of the present invention, the stored content management means (corresponding to the stored content storage step) re-organizes the media content stored in the data recording medium according to the period of storage of the content.
10   Therefore, a larger amount of media content can be stored in the data recording medium.

<center>BRIEF DESCRIPTION OF THE DRAWINGS</center>

FIG. 1 is a block diagram showing a data processing method according to a first embodiment of the present invention;

15   FIG. 2 is a diagram showing the structure of context description data according to the first embodiment;

FIG. 3 shows a portion of one example of Document Type Definition (DTD) used for describing the context description data in a computer according to the first embodiment through use of XML, as well as a portion of one example of context description data described through use of DTD according to the first embodiment;

FIGS. 4-9 show continued portions of the context description data of the example shown in FIG. 3;

25   FIG. 10 shows a portion of one example of the XML document

<center>16</center>

formed by addition of representative data to the context description data shown in FIGS. 3 through 9, as well as a portion of one example of DTD described in Extensible Markup Language (XML) for describing the context description data in a computer;

5   FIGS. 11-21 show continued portions of the context description data shown in FIG. 10;

FIG. 22 is a descriptive view for describing a method of assigning the degree of importance according to the first embodiment;

10   FIG. 23 is a flowchart showing processing relating to the selection step according to the first embodiment;

FIG. 24 is a block diagram showing the configuration of the extraction step according to the first embodiment;

FIG. 25 is a flowchart showing processing effected by

15   demultiplexd means in the extraction step according to the first embodiment;

FIG. 26 is a flowchart showing processing effected by video skimming means in the extraction step according to the first embodiment;

20   FIG. 27 is a schematic representation showing the configuration of an MPEG-1 video stream;

FIG. 28 is a flowchart showing processing effected by audio skimming means in the extraction step according to the first embodiment;

25   FIG. 29 is a schematic representation showing the

configuration of AAUs of the MPEG-1 audio stream;

FIG. 30 is a block diagram showing an application of the media processing method according to the first embodiment;

FIG. 31 is a descriptive view showing processing of the degree of importance according to a second embodiment of the present invention;

FIG. 32 is a flowchart showing processing relating to the selection step according to the second embodiment;

FIG. 33 is a flowchart showing processing relating to the selection step according to a third embodiment of the present invention;

FIG. 34 is a descriptive view for describing a method of assigning the degree of importance according to a fourth embodiment of the present invention;

FIG. 35 is a flowchart showing processing relating to the selection step according to the fourth embodiment;

FIG. 36 is a block diagram showing a media processing method according to a fifth embodiment of the present invention;

FIG. 37 is a diagram showing the structure of structure description data according to the fifth embodiment;

FIG. 38 is a diagram showing the structure of context description data according to the fifth embodiment;

FIG. 39 shows one example of Document Type Definition (DTD) used for describing the structure description data in a computer according to the fifth embodiment through use of XML, as well

as one example of a XML document, according to the fifth embodiment;

FIG. 40 shows a first half of one example of Document Type Definition (DTD) used for describing the context description data in a computer according to the fifth embodiment through use of XML, as well as a first half of one example of an XML document, according to the fifth embodiment;

FIGS. 41-45 show continued portions of the context description data shown in Fig. 40;

FIG. 46 shows one example of an output in the selection step according to the fifth embodiment;

FIG. 47 is a block diagram showing the extraction step according to the fifth embodiment;

FIG. 48 is a flowchart showing processing effected by interface means in the extraction step according to the fifth embodiment;

FIG. 49 shows one example of a result produced when the interface means provided in the extraction step converts the output in the selection step according to the fifth embodiment;

FIG. 50 is a flowchart showing processing effected by demultiplex means in the extraction step according to the fifth embodiment;

FIG. 51 is a flowchart showing processing effected by video skimming means in the extraction step according to the fifth embodiment;

19

FIG. 52 is a flowchart showing processing effected by audio skimming means in the extraction step according to the fifth embodiment;

FIG. 53 is another flowchart showing processing effected by video skimming means in the extraction step according to the fifth embodiment;

FIG. 54 is a block diagram showing a data processing method according to a sixth embodiment of the present invention;

FIG. 55 is a block diagram showing the formation step and the delivery step according to the sixth embodiment;

FIG. 56 is a block diagram showing a media processing method according to a seventh embodiment of the present invention;

FIG. 57 is a diagram showing the structure of context description data according to the fifth embodiment;

FIG. 58 shows a portion of one example of Document Type Definition (DTD) used for describing context description data in a computer according to a seventh embodiment through use of XML, as well as a portion of one example of context description data described through use of XML, according to the seventh embodiment;

FIGS. 59-66 show continued portions of the context description data shown in FIG. 58;

FIG. 67 shows a portion of one example of the XML document formed by addition of representative data to the context description data shown in FIGS. 58 through 66, as well as a portion

of one example of DTD described in XML for describing the context description data in a computer;

FIGS. 68-80 show continued portions of the context description data shown in FIG. 67;

FIG. 81 is a flowchart showing processing pertaining to the selection step according to the seventh embodiment;

FIG. 82 is a block diagram showing an application of the media processing method according to the seventh embodiment;

FIG. 83 is a flowchart showing processing pertaining to the selection step according to an eighth embodiment of the present invention;

FIG. 84 is a flowchart showing processing pertaining to the selection step according to an ninth embodiment of the present invention;

FIG. 85 is a flowchart showing processing pertaining to the selection step according to a tenth embodiment of the present invention;

FIG. 86 is a block diagram showing a data processing method according to a twelfth embodiment of the present invention;

FIG. 87 is a diagram showing the structure of context description data according to the twelfth embodiment;

FIG. 88 shows a portion of one example of Document Type Definition (DTD) used for describing context description data in a computer according to the fifth embodiment through use of XML, as well as a portion of one example of an XML document,

21

according to the fifth embodiment;

FIGS. 89-96 show continued portions of the context description data shown in FIG. 88;

5   FIG. 97 is a block diagram showing a data processing method according to a thirteenth embodiment of the present invention;

FIG. 98 is a block diagram showing a data processing method according to a fourteenth embodiment of the present invention;

FIG. 99 is a block diagram showing a data processing method

10   according to a fifteenth embodiment of the present invention;

FIG. 100 is a block diagram showing a data processing method according to a sixteenth embodiment of the present invention;

FIG. 101 is a block diagram showing a data processing method according to a seventeenth embodiment of the present invention;

15   FIG. 102 is a descriptive view showing channels and layers;

FIG. 103 shows a portion of one example of Document Type Definition (DTD) used for describing structure description data through use of XML, as well as a portion of one example of the structure description data described in DTD;

20   FIG. 104 shows a continued portion of the structure description data shown in FIG. 103;

FIG. 105 is a flowchart showing processing pertaining to the determination step in example 1 according to a seventeenth embodiment of the present invention;

25   FIG. 106 is a flowchart showing determination processing

to be performed, in response to a user request, in the determination step of example 1 according to the seventeenth embodiment;

FIG. 107 is a flowchart showing determination processing pertaining to video data in the determination step of example 1 according to the seventeenth embodiment;

FIG. 108 is a flowchart showing determination processing pertaining to sound data in the determination step of example 1 according to the seventeenth embodiment;

FIG. 109 is a flowchart showing a first half of processing pertaining to the determination step in example 2 according to a seventeenth embodiment of the present invention;

FIG. 110 is a flowchart showing a second half of processing pertaining to the determination step in example 2 according to a seventeenth embodiment of the present invention;

FIG. 111 is a flowchart showing processing pertaining to the determination step in example 3 according to a seventeenth embodiment of the present invention;

FIG. 112 is a flowchart showing determination processing pertaining to video data in the determination step of example 3 according to the seventeenth embodiment;

FIG. 113 is a flowchart showing determination processing pertaining to sound data in the determination step of example 3 according to the seventeenth embodiment;

FIG. 114 is a flowchart showing a first half of processing

23

pertaining to the determination step in example 4 according to a seventeenth embodiment of the present invention;

FIG. 115 is a flowchart showing a second half of processing pertaining to the determination step in example 4 according to a seventeenth embodiment of the present invention;

FIG. 116 is a flowchart showing determination processing to be performed, in response to a user request, in the determination step of example 4 according to the seventeenth embodiment;

FIG. 117 is a flowchart showing determination processing pertaining to video data in the determination step of example 4 according to the seventeenth embodiment;

FIG. 118 is a flowchart showing determination processing pertaining to sound data in the determination step of example 4 according to the seventeenth embodiment;

FIG. 119 is a flowchart showing a first half of processing pertaining to the determination step in example 5 according to a seventeenth embodiment of the present invention;

FIG. 120 is a flowchart showing a second half of processing pertaining to the determination step in example 5 according to a seventeenth embodiment of the present invention;

FIG. 121 is a flowchart showing determination processing to be performed, in response to a user request, in the determination step of example 5 according to the seventeenth embodiment;

24

FIG. 122 is a block diagram showing a data processing method according to a eighteenth embodiment of the present invention;

FIG. 123 is a block diagram showing a data processing method according to a nineteenth embodiment of the present invention;

FIG. 124 is a block diagram showing a data processing method according to a twentieth embodiment of the present invention;

FIG. 125 is a block diagram showing a data processing method according to a twenty-first embodiment of the present invention;

FIG. 126 is a block diagram showing a data processing method according to a twenty-second embodiment of the present invention;

FIG. 127 shows one example of a DTD into which context description data and structure description data are to be merged, as well as one example of an XML document;

FIGS. 128-132 continued portions of the XML document shown in FIG. 127;

FIG. 133 is an illustration showing the structure of context description data according to an eleventh embodiment of the present invention;

FIG. 134 is an illustration showing a viewpoint employed in the eleventh embodiment;

FIG. 135 is an illustration showing the degree of importance according to the eleventh embodiment;

FIG. 136 is an example of DTD used for describing the context description data of the eleventh embodiment through use of XML to be used in expressing the context description data in a computer,

and an example of a portion of the context description data described in XML;

FIGS. 137 to 163 show continued portions of the context description data shown in FIG. 136;

FIG. 164 is another example of DTD used for describing the context description data of the eleventh embodiment through use of XML to be used in expressing the context description data in a computer, and an example of a portion of the context description data described in XML;

FIGS. 165 to 196 show continued portions of the context description data shown in FIG. 164;

FIG. 197 is an illustration showing another structure of context description data according to an eleventh embodiment of the present invention;

FIG. 198 is an example of DTD used for describing the context description data (corresponding to Fig. 197) of the eleventh embodiment through use of XML to be used in expressing the context description data in a computer, and an example of a portion of the context description data described in XML;

FIGS. 199 to 222 show continued portions of the context description data shown in FIG. 164;

FIG. 223 is another example of DTD used for describing the context description data (corresponding to Fig. 197) of the eleventh embodiment through use of XML to be used in expressing the context description data in a computer, and an example of

26

a portion of the context description data described in XML;

FIGS. 224 to 252 show continued portions of the context description data shown in FIG. 164;

Fig. 253 is a diagram showing a link of a view point in the program expression of the context description data;

Figs. 254 to 256 are diagrams showing a view point table and links of view points in the program expression of the context description data;

Figs. 257 to 260 are diagrams showing the data structure of the context description data formed by a data structure portion and an attribute portion;

Figs. 261 to 263 are diagrams showing the data structure of the second context description data formed by the first embodiment of the context description data converting method;

Fig. 264 is a diagram showing an example the <segments> and (keywords, priorities) of the original context description data;

Fig. 265 is a diagram showing the data structure of the second context description data which is formed by converting the original context description data shown in Fig. 264 by the first embodiment of the context description data converting method;

Fig. 266 is a diagram showing the data structure of the second context description data formed by the second embodiment of the context description data converting method;

27

Fig. 267 is a diagram showing the data structure of the second context description data which is formed by converting the original context description data shown in Fig. 264 by the second embodiment of the context description data converting method;

Fig. 268 is a diagram showing the data structure of the second context description data formed by the third embodiment of the context description data converting method;

Fig. 269 is a diagram showing another data structure of the second context description data formed by the third embodiment of the context description data converting method; and

Fig. 270 is a diagram showing the data structure of the second context description data which is formed by converting the original context description data shown in Fig. 264 by the third embodiment of the context description data converting method.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Embodiments of the present invention will be described hereinbelow by reference to the accompanying drawings.

[First Embodiment]

A first embodiment of the present invention will now be described. In the present embodiment, a motion picture of MPEG-1 system stream is taken as media content. In this case, a media segment corresponds to a single scene cut, and a score represents

28

the objective degree of contextual importance of a scene of interest.

FIG. 1 is a block diagram showing a data processing method according to the first embodiment of the present invention. In FIG. 1, reference numeral 101 designates the selection step; and 102 designates an extraction step. In the selection step 101, a scene of media content is selected from context description data, and the start time and the end time of the scene are output. In the extraction step 102, data pertaining to a segment of media content defined by the start time and the end time output in the selection step 101 are extracted.

FIG. 2 shows the configuration of the context description data according to the first embodiment. In the present embodiment, the context is described according to a tree structure. Elements within the tree structure are arranged in chronological sequence from left to right. In FIG. 2, the root of the tree designated <contents> represents a single portion of content, and the title of the content is assigned to the root as an attribute.

Children of <program> are designated by <section>. Priority representing the degree of contextual importance of a scene of interest is appended to the element <section> as an attribute. The degree of importance assumes an integral value ranging from 1 to 5, where 1 designates the least degree of importance and 5 designates the greatest degree of importance.

29

Children of <section> are designated by <section> or . Here, an element <section> per se can be taken as a child of another child <section>. However, a single element <section> cannot have a mixture of children <section> and

5   children .

An element represents a single scene cut and is assigned a priority identical with that assigned to its parent <section>. Attributes appended to are "start" representing the start time and "end" representing the end time.

10   Scenes may be cut through use of commercially-available software or software available over a network. Alternatively, scenes may be cut manually. Although in the present embodiment time information is expressed in terms of the start time and the end time of a scene cut, a similar result is realized when time

15   information is expressed in terms of the start time of the scene of interest and the duration of the scene of interest. In this case, the end time of the scene of interest is obtained by addition of the duration to the start time.

In [KIK16] the case of a story such as a movie, chapters,

20   sections, and paragraphs of the story can be described on the basis of the context description data, through use of elements <section> within a multilayer hierarchical stratum. In another example, when a baseball game is described, elements <section> in the highest hierarchical level may be used for describing

25   innings, and their children <section> may be used for describing

30

half innings. Further, second-generation descendant <section> of the elements <section> may be used for describing at-bats of respective batters, and third-generation descendant <section> of the elements <section> are also used for describing

5   each pitch, a time period between pitches, and batting results.

The context description data having such a configuration may be expressed in a computer through use of, e.g., Extensible Markup Language (XML). XML is a data description language whose standardization is pursued by the World Wide Web Consortium.

10   Recommendations Ver. 1.0 were submitted on February 10, 1998. Specifications of XML Ver. 1.0 can be acquired from http://www.w3.org/TR/1998/REC-xml-19980210. FIGS. 3 through 9 show one example of Document Type Definition (DTD) used for describing the context description data according to the present

15   embodiment through use of XML, and one example of context description data described through use of DTD. FIGS. 10 through 19 show one example of context description data prepared by addition of representative data (dominant-data) of a media segment, such as a representative image (i.e., video data) and

20   a keyword (audio data), to the context description data shown in FIGS. 3 through 9, and a DTD used for describing the context description data through use of XML.

Processing relating to the selection step 101 will now be described. Processing pertaining to the selection step 101

25   closely relates to the format of context description data and

31

a method of assigning a score to contents of a context of each scene. In the present embodiment, processing pertaining to the selection step 101 is effected by focusing on only elements <section> having children , as shown in FIG. 22 (steps

5   S1, S4, and S5 shown in FIG. 23). An element <section> whose priority exceeds a certain threshold value is selected (step S2 shown in FIG. 23), and the start time and end time of the thus-selected element <section> are output (step S3 shown in FIG. 23). The priority assigned to the element <section> having

10 children corresponds to the degree of importance shared among all the elements <section>, each of which has children , within the content. More specifically, the degree of importance shared among the elements <section> enclosed by a dotted line shown in FIG. 22 is set as priority. Priority

15 assigned to elements <section> and other than the foregoing elements <section> is set arbitrarily. The degree of importance are not necessarily set so as to assume unique values, and the same degree of importance may be assigned to different elements. FIG. 23 is a flowchart showing processing

20 relating to the selection step 101 according to the first embodiment. With regard to the thus-selected element <section>, the start time and end time of scene expressed by the element <section> are determined from elements , which are children of the thus-selected element <section>. The

25 thus-determined start time and end time are output.

32

Although in the present embodiment selection is effected by focusing on the elements <section>, each of which has children , selection may be effected by focusing on elements . In this case, priority corresponds to the degree of importance shared among all the elements within the content. Alternatively, selection may be effected by focusing on elements <section> of the same hierarchical level from among the elements <section> of higher hierarchical levels having no children . More specifically, selection may be effected by focusing on the elements <section> in the same path number, which is counted from a given parent <contents> or a given child .

Processing relating to the extraction step 102 will now be described by reference to FIG. 24. FIG. 24 is a block diagram showing the extraction step 102 according to the first embodiment. As shown in FIG. 24, the extraction step 102 according to the first embodiment is realized by demultiplex means 601, video skimming means 602, and audio skimming means 603. In the present embodiment, an MPEG-1 system stream is taken as media content. The MPEG-1 system stream is formed by multiplexing a video stream and an audio stream into a single stream. The demultiplex means 601 separates the video stream and the audio stream from the multiplexed system stream. The video skimming means 602 receives the thus-separated video stream and a segment selected in the selection step 101, and from the received video stream

33

outputs only data pertaining to the thus-selected segment. The audio skimming means 603 receives the separated audio stream and the segment selected in the selection step 101, and from the received audio stream outputs only data pertaining to the

5   selected segment.

The processing performed by the demultiplex means 601 will now be described by reference to the accompanying drawings. FIG. 25 is a flowchart relating to processing effected by the demultiplex means 610. The method of multiplexing the MPEG-1

10   system stream is standardized under International Standard ISO/IEC IS 11172-1. A video stream and an audio stream are multiplexed into packets by means of dividing the video and audio streams into streams of appropriate length called packets and by appending additional information, such as a header, to each

15   of the packets. A plurality of video streams and a plurality of audio streams may also be multiplexed into a single signal in the same manner. In the header of each packet, there are described a stream ID for identifying a packet as a video stream or an audio stream, and a time stamp for bringing video data

20   into synchronization with audio data. The stream ID is not limited to use for identifying a packet as a video stream or an audio stream. When a plurality of video streams are multiplexed, the stream ID can be used for identifying, from a plurality of video streams, the video stream to which a packet

25   of interest belongs. Similarly, when a plurality of audio

streams are multiplexed, the stream ID can be used for identifying, from a plurality of audio streams, the audio stream which a packet of interest belongs. In the MPEG-1 system, a plurality of packets are bundled into a single pack, and to the pack is appended,

5 as a header, a multiplex rate and additional information for use as a reference time used for effecting synchronous playback. Further, additional information relating to the number of multiplexed video and audio streams is appended, as a system header, to the head pack. The demultiplex means 601 reads the

10 number of multiplexed video and audio streams from the system header of the head pack (S1 and S2) and ensures data locations for storing data sets of the respective streams (S3 and S4). Subsequently, the demultiplex means 601 examines the stream ID of each of the packets and writes the data included in the packet

15 into the data location where the stream specified by the stream ID is stored (S5 and S6). All the packets are subjected to the foregoing processing (S8, S9, and S10). After all the packets have been subjected to the processing, the video streams are output to the video skimming means 602 on a per-stream basis,

20 and the audio streams are output to the audio skimming means 603 in the same manner (S11).

The operation of the video skimming means 602 will be described hereinbelow. FIG. 26 is a flowchart relating to processing effected by the video skimming means 602. The MPEG-1

25 system stream is standardized under International Standard

35

ISO/IEC IS 11172-2. As shown in FIG. 27, the video stream comprises a sequence layer, a group-of-pictures (GOP) layer, a picture layer, a slice layer, a macro block layer, and a block layer. Random access is made on the basis of GOP layer, which

5   is the minimum unit, and each layer included in the picture layer corresponds to a single frame. The video skimming means 602 processes data on a per-GOP basis. A counter C for counting the number of output frame is initialized to 0 (S3). First, the video skimming means 602 acknowledges that the header of

10   the video stream corresponds to the header of the sequence layer (S2 and S4) and stores data included in the header (S5). Subsequently, video skimming means outputs the data. The header of the sequence layer may appear during subsequent processes. The value of the header is not allowed to be changed unless the

15   value is relevant to a quantization matrix. Therefore, every time the sequence header is input, the value of the input header is compared with the value of the stored header (S8 and S14). If the input header differs from the stored header in terms of a value other than the value relevant to the quantization matrix,

20   the input header is considered an error (S15). Subsequently, the video skimming means 602 detects the header of the GOP layer of the input data (S9). Data pertaining to a time code are described in the header of the GOP layer (S10), and the time code describes the period of time which has elapsed from the

25   head of the sequence. The video skimming means 602 compares

36

the time code with the segment output in the selection step 101 (S1) (S11). If the time code is determined not to be included in the segment, the video skimming means 602 discards all the data sets appearing before the next GOP layer of the sequence

5   layer. In contrast, if the time code is included in the selected segment, the video skimming means 602 outputs all the data sets appearing before the next GOP layer of the sequence layer (S13). In order to ensure continuity the data sets which have already been output, and the data sets currently being output, the time

10   code of the GOP layer must be changed (S12). A value to which the time code of the GOP layer is to be changed is computed through use of the value of the counter C. The counter C retains the number of frames which have already been output. In accordance with Eq. 1, the time Tv at which the header frame of the GOP

15   layer to be currently output is displayed is computed from C, as well as from a picture rate "pr" which is described in the sequence header and represents the number of frames to be displayed per second.

$$Tv = \frac{C}{pr} \qquad \dots (1)$$

20   "Tv" designates a value in units of 1/pr sec, and hence the value of Tv is converted in accordance with the format of the time code of the MPEG-1. the thus-converted value is set in the time code of the GOP layer which is to be output at this time. When the data pertaining to the GOP layer are output,

37

the number of output picture layers is added to the value of

the counter C. The foregoing processing is performed repeatedly

until the end of the video stream (S7 and S16). In a case where

the demultiplex means 601 outputs a plurality of video streams,

5 the processing is performed for each of the video streams.

Processing of the audio skimming means 603 will now be

described. FIG. 28 is a flowchart relating to processing

effected by the audio skimming means 603. The MPEG-1 audio stream

is standardized under International Standard ISO/IEC IS 11172-3.

10 The audio stream is formed from a series of frames called audio

access units (AAUs). FIG. 29 shows the structure of an AAU.

The AAU is the minimum unit at which audio data can be decoded

independently and comprises a given number of sampled data sets

Sn. The playback time of a single AAU can be computed from a

15 bit rate "br" representing the transmission rate; a sampling

frequency Fs; and the number of bits, L, of the AAU. First,

the header of the AAU included in the audio stream is detected

(S2 and S5), thereby obtaining the number of bits, L, of a single

AAU. Further, the bit rate "br" and the sampling frequency Fs

20 are described in the header of the AAU. The number of samples,

Sn, of a single AAU is calculated in accordance with Eq. 2.

$$Sn = \frac{L \times Fs}{br} \quad \ldots \quad (2)$$

The playback time Tu of a single AUU is computed in accordance

with Eq. 3.

$$Tu = \frac{Sn}{Fs} = \frac{L}{br} \qquad \dots (3)$$

So long as the value of Tu is computed, the time which has elapsed from the head of the stream can be obtained by counting the number of AAUs. The audio skimming means 603 counts the number of AAUs which have already appeared and calculates the time which has elapsed from the head of the stream (S7). The thus-calculated time is compared with the segment output in the selection step 101 (S8). If the time at which the AAU appears is included in the selected segment, the audio skimming means 603 outputs all the data sets relating to that AAU (S9). In contrast, if the time at which the AAU appears is not included in the selected segment, the audio skimming means 603 discards the data sets pertaining to the AAU. The foregoing processing is performed repeatedly until the end of the audio stream (S6 and S11). When the demultiplex means 601 outputs a plurality of audio streams, each of the audio streams is subjected to the previously-described processing.

As shown in FIG. 30, the video stream output from the extraction step 102 is input to video playback means, and the audio stream output from the extraction step 102 is input to audio playback means. The video stream and the audio stream are played back synchronously, thereby enabling playback of a synopsis or a highlight scene of media content. Further, the thus-produced video and audio streams are multiplexed, thereby

enabling preparation of an MPEG-1 system stream relating to a synopsis of the media content or a collection of highlight scenes of the same.

[Second Embodiment]

5   A second embodiment of the present invention will now be described. The second embodiment differs from the first embodiment only in terms of processing relating to the selection step.

Processing relating to the selection step 101 according 10 to the second embodiment will now be described by reference to the drawings. In the selection step 101 according to the second embodiment, the priority values assigned to all the elements ranging from <section> of the highest hierarchical level to leaves are utilized. The priority assigned to each 15 of the elements <section> and represents the objective degree of contextual importance. Processing relating to the selection step 101 will now be described by reference to FIG. 31. In FIG. 31, reference numeral 1301 designates one of elements <section> of the highest hierarchical level included in the 20 context description data; 1302 designates a child element <section> of the element <section> 1301; 1303 designates a child element <section> of the element <section> 1302; and 1304 designates a child element of the element <section> 1303. In the selection step 101 according to the second 25 embodiment, an arithmetic mean of all the priority values

40

assigned to the path extending from the leaf to its ancestor <section> of the highest hierarchical level is calculated. When arithmetic means of the path exceeds a threshold value, the element is selected. In the example shown in FIG. 28, an arithmetic mean "pa" of the attributes of elements: 1304, <section> 1303, <section> 1302, and <section> 1301; i.e., the arithmetic mean of their attribute priority values p4, p3, p2, and p1, is calculated. The arithmetic mean "pa" is calculated in accordance with Eq. 4.

$$pa = \frac{p1 + p2 + p3 + p4}{4} \qquad \ldots \ (4)$$

.The thus-calculated "pa" is compared with the threshold value (S1 and S2). If "pa" exceeds the threshold value, 1304 is selected (S3), and the attribute values relating to "start" and "end" of 1304 are output as the start time and end time of the selected scene (S4). All the elements are subjected to the foregoing processing (S1 and S6). FIG. 32 is a flowchart showing processing relating to the selection step 101 according to the second embodiment.

In the second embodiment, an arithmetic mean of the priority value assigned to the of the lowest hierarchical level up to the priority value assigned to its ancestor <section> of the highest hierarchical level is calculated, and the leaf is selected on the basis of the thus-calculated arithmetic mean. Alternatively, there may be calculated an

41

arithmetic mean of the priority values assigned to the element <section> having a child up to the priority value assigned to its ancestor <section> of the highest hierarchical level, and the element <section> having the child may be selected by comparing the thus-calculated arithmetic mean with the threshold value. Similarly, in another hierarchical stratum, an arithmetic mean of the priority value assigned to an element <section> up to the priority value assigned to its ancestor <section> of the highest hierarchical level is calculated, and the thus-calculated arithmetic mean is compared with the threshold value, whereby the element <section> in the hierarchical stratum may be selected.

[Third Embodiment]

A third embodiment of the present invention will now be described. The third embodiment differs from the first embodiment only in terms of the processing relating to the selection step.

The processing relating to the selection step 101 according to the third embodiment will be described by reference to the drawings. As in the case of the processing described in connection with the first embodiment, in the selection step 101 according to the third embodiment, selection is effected by focusing on only the elements <section>, each of which has a child . In the third embodiment, there is set a threshold value with regard to the sum of the duration periods

42

of all the scenes to be selected. More specifically, elements <section> are selected in decreasing order of priority value, until the sum of the duration periods of the elements <section> that have been selected so far is maximized but remains smaller than the threshold value. FIG. 33 is a flowchart of processing pertaining to the selection step 101 according to the third embodiment. A collection of elements <section>, each of which has children , is taken as a set Ω (S1). The elements <section> of the set Ω are sorted in descending order of attributes priority (S2). The element <section> having the highest priority value is selected from the set Ω (S4 and S5), and the thus-selected element <selection> is eliminated from the set Ω. The start time and end time of the thus-selected element <section> are obtained by examination of all the children of the element <section>, and a duration of the element <section> is calculated (S6). The sum of the duration periods of the elements <section> which have been selected so far is calculated (S7). If the sum exceeds the threshold value, processing is completed (S8). If the sum is lower than the threshold value, the start time and the end time of the element <section> selected this time are output (S9). Processing then returns to a step in which the element <section> having the highest priority value is selected from the set Ω. The above-described processing is repeated until the sum of duration periods of the

43

selected elements <section> exceeds the threshold value or the set Ω becomes empty (S4 and S8).

In the third embodiment, selection is effected by focusing on the element <section> having children . However, selection may be effected by focusing on elements in place of the elements <section>. In this case, a priority value corresponds to the degree of importance shared among all the elements within the media content. Further, selection may be effected by focusing on the elements <section> having no children within the same hierarchical level. More specifically, selection may be effected by focusing on the elements <section> located in the same path, which is counted from the ancestor <contents> or a leaf .

As in the case of the second embodiment, the priority values assigned to the respective elements <section> and are taken as the objective degree of contextual importance, and the arithmetic mean "pa" of all the priority values assigned to the element up to its ancestor <section> of the highest hierarchical level is calculated. Elements <section>, each having children , or elements are selected in descending order of "pa" until the sum of duration periods is maximized but remains smaller than the threshold value. Even in this case, the same advantageous result as that yielded in the second embodiment is achieved.

[Fourth Embodiment]

44

A fourth embodiment of the present invention will now be described. The fourth embodiment differs from the first embodiment only in terms of the processing relating to the selection step.

5   Processing relating to the selection step 101 according to the fourth embodiment will now be described by reference to the drawings. As in the case of the selection performed in the selection step 101 in the first embodiment, selection relating to the selection step 101 in the fourth embodiment is effected

10   by focusing on an element and an element <section> having children . As in the case of the third embodiment, a threshold value is set with regard to the sum of duration periods of all scenes to be selected in the present embodiment. As in the case of the first embodiment, the priority value assigned

15   to the element <section> having children corresponds to the degree of importance shared among all the elements <section>, each of which has children , within the media content. More specifically, the priority value is taken as a degree of importance shared among the elements <section> enclosed

20   by a dotted line shown in FIG. 34. Further, the priority value assigned to the element corresponds to the degree of importance shared among the elements sharing the same parent element <section>; that is, the degree of importance shared among the elements enclosed by one of the dashed

25   lines shown in FIG. 34.

45

FIG. 35 is a flowchart showing processing relating to the selection step 101 according to the fourth embodiment. A collection of elements <section>, each of which has children , is taken as set Ω (S1). The elements <section> within the set Ω are sorted in descending order of priority (S2). Subsequently, the element <section> having the highest priority value is selected from the set Ω (S3, S4, and S5). If a plurality of elements <section> have the highest priority value, all the elements are selected. The thus-selected elements <section> are taken as elements of another set Ω' and are eliminated from the set Ω. The start time, the end time, and a duration of a scene represented by the thus-selected element <section> are obtained and stored in advance by examination of the children of the element <section> (S6). If the plurality of elements <section> are selected, the start time, the end time, and the duration of each of the scenes represented by the respective elements are obtained and stored in advance. The sum of duration periods of the elements <section> of the set Ω' is obtained (S7 and S8). The sum is compared with a threshold value (S9). If the sum of duration periods is equal to the threshold value, all the data sets which pertain to the start time and the end time and have been stored so far are output, and processing is terminated (S10). In contrast, if the sum of duration periods is lower than the threshold value, processing

46

again returns to the selection of an element <section> from the set Ω (S4 and S5). If the set Ω is empty, all the data sets pertaining to the start time and the end time that are stored are output, and processing is terminated (S4). If the sum of duration periods exceeds the threshold value, the following processing is performed. Specifically, the element <section> having the minimum priority is selected from the set Ω' (S11). At this time, if a plurality of elements <section> have the minimum priority, all the elements are selected. Of the children of the thus-selected elements <section>, the children having the minimum priority are deleted (S12). The start time, the end time, and the duration of the element <section> corresponding to the thus-eliminated children are changed (S13). As a result of deletion of the elements , scenes may be interrupted. In such a case, for each of the scenes, which have been interrupted, the start time, the end time, and a duration are stored. Further, if, as a result of deletion of the children , all the children of an element <section> are deleted, the element <section> is deleted from the set Ω'. If the plurality of elements <section> are selected, all the elements are subjected to the previously-described processing. As a result of deletion of the children , the duration of the element <section> from which the children have been deleted becomes shorter, in turn reducing

47

the sum of duration periods. Such deletion processing is performed repeatedly until the sum of duration periods of the elements of the set Ω' becomes lower than the threshold value. When the sum of the duration periods of the elements of the set

5 Ω' becomes lower than the threshold value (S14), all the data sets which pertain to the start time and the end time and have been stored are output, and processing is terminated (S15).

Although in the fourth embodiment selection is effected by focusing on the elements <section>, each of which has children

10 , or elements , selection may also be effected by focusing on an element <section> and its children <section> or an element <section> and its children . Even in such a case, the same advantageous result as that yielded by the fourth embodiment is achieved.

15 With regard to deletion of the elements effected when the sum of duration periods exceeds the threshold value, in the present embodiment the elements <section> are deleted in ascending sequence of priority from the lowest priority. However, a threshold value may be set for the priority of elements

20 <section>, and the children having the minimum priority may be deleted from all the elements <section> which are lower than the threshold value. Alternatively, another threshold value may be set for the priority of elements , and elements whose priority is lower than the threshold

48

value may be deleted.

[Fifth Embodiment]

A fifth embodiment of the present invention will now be described by reference to the accompanying drawings. In the present embodiment, a motion picture of MPEG-1 format is taken as media content. In this case, a media segment corresponds to a single scene cut, and a score corresponds to the objective degree of contextual importance of a scene of interest.

FIG. 36 is a block diagram showing a media processing method according to the fifth embodiment of the present invention. In FIG. 36, reference numeral 1801 designates a selection step; 1802 designates an extraction step; 1803 designates a formation step; 1804 designates a delivery step; and 1805 designates a database. In the selection step 1801, a scene of media content is selected from context description data, and there are output data pertaining to the start time and the end time of the thus-selected scene, as well as data representing a file where the data are stored. In the extraction step 1802, there are received the data sets representing the start time and the end time of the scene and the data sets representing the file output in the selection step 1801. Further, in the extraction step 1802, by reference to the structure description data, data pertaining to the segment defined by the start time and the end time output in the selection step 1801 are extracted from the file of media content. In the formation step 1803, the data

49

output in the extraction step 1802 are multiplexed, thus
configuring a system stream of MPEG-1 format. In the delivery
step 1804, the system stream of MPEG-1 format prepared in the
formation step 1803 is delivered over a line. Reference numeral

5   1805 designates a database where media content, structure
description data thereof, and context description data are
stored.

FIG. 37 shows the configuration of the structure description
data according to the fifth embodiment. In the present

10   embodiment, the physical contents of the data are described in
a tree structure. With regard to the nature of storage of media
content in the database 1805, a single piece of media content
is not necessarily stored in the form of a single file. In some
cases, a single piece of media content may be stored in a plurality

15   of separate files. The root of the tree structure of structure
description data is depicted as <contents> and represents a
single piece of content. The title of a corresponding piece
of content is appended to the root <contents> as an attribute.
A children of <contents> corresponds to <mediaobject>, which

20   represents a file where the media content is stored. The child
<mediaobject> is appended, as an attribute, to a link "locator"
representing a link to the file where the media content is stored
and an identifier ID representing a link to context description
data. In a case where media content is constituted of a plurality

25   of files, "seq" is appended to the element <mediaobject> as an

attribute for representing the sequence of a file of interest within the media content.

FIG. 38 shows the configuration of the context description data according to the fifth embodiment. The context description data of the present embodiment corresponds to the context description data of the first embodiment appended with a link to the element <mediaobject> of the structure description data. More specifically, the root <contents> of the context description data has a child <mediaobject>, and the element <mediaobject> has a child <section>. Elements <section> and are identical with those used in the first embodiment. The element <mediaobject> of the structure description data is associated with the element <mediaobject> of the context description data. Scenes of the media content described by means of children of the element <mediaobject> of the context description data are stored in a file designated by the element <mediaobject> of the structure description data having the attribute ID of the same value. Further, time information "start" and "end" assigned to an element "segment" sets the time which has elapsed from the head of each file. Specifically, in a case where a single piece of media content comprises a plurality of files, the time at the head of each file corresponds to 0, and the start time of each scene is represented by the time which has elapsed from the head of the file to a scene of interest.

The structure description data and the context description

51

data may be expressed in a computer through use of, e.g., Extensible Markup Language (XML). FIG. 39 shows one example of Document Type Definition (DTD) used for describing the structure description data shown in FIG. 37 through use of XML,

5 as well as one example of structure description data described through use of DTD. FIGS. 40 through 45 show DTD used for describing the context description data shown in FIG. 38 through use of XML and one example of the context description data described by DTD.

10 Processing relating to the selection step 1801 will now be described. In the selection step 1801, any one of the methods described in connection with the first through fourth embodiments is adopted as a method of selecting a scene. A link to <object> of structure description data is eventually output

15 simultaneously with output of the start time and the end time of a selected scene. FIG. 46 shows one example of data output from the selection step 1801 in a case where the structure description data are described in the form of an XML document through use of the DTD shown in FIG. 39 and where the context

20 description data are described in the form of an XML document through use of the DTD shown in FIGS. 40 and 45. In FIG. 46, "id" is followed by an ID of an element <mediaobject> of structure description data; "start" is followed by the start time; and "end" is followed by the end time.

25 Processing relating to the extraction step 1802 will now

52

be described. FIG. 47 is a block diagram showing the extraction

step 1802 according to the fifth embodiment. In FIG. 47, the

extraction step 1802 according to the fifth embodiment is

embodied by interface means 2401, demultiplex means 2402, video

5   skimming means 2403, and audio skimming means 2404. The

interface means 2401 receives structure description data and

a segment output in the selection step 1801, extracts a file

of media content from the database 1805, outputs the

thus-extracted file to the demultiplex means 2402, and outputs

10   to the video skimming means 2403 and the audio skimming means

2404 the start time and end time of the segment output in the

selection step 1801. Media content of the present embodiment

corresponds to a system stream of MPEG-1 format into which a

video stream and an audio stream are multiplexed. Accordingly,

15   the demultiplex means 2402 separates the system stream of MPEG-1

format into the video stream and the audio stream. The

thus-separated video stream and the segment output from the

interface means 2401 are input to the video skimming means 2403.

From the input video stream, the video skimming means 2403 outputs

20   only the data pertaining to the selected segment. Similarly,

the audio stream and the segment output in the selection step

2402 are input to the audio skimming means 2404. From among

the input audio stream, the audio skimming means 2402 outputs

only the data pertaining to the selected segment.

25   Processing relating to the interface means 2401 will now

53

be described. FIG. 48 is a flowchart showing processing effected by the interface means 2401. Structure description data pertaining to corresponding content and the segment output in the selection step 1801, as shown in FIG. 46, are input to the interface means 2401. Chronological order of files is acquired from the attribute "id" assigned to the element <mediaobject> of the structure description data, and hence the segments output in the selection step 1801 are sorted in chronological sequence and in order of "id" (S1). Further, the segments are converted into data such as those shown in FIG. 49. The same files are collected and arranged in sequence of start time. Subsequently, the interface means 2401 subjects the data sets shown in FIG. 49 to the following processing in sequence from top to bottom. First, the interface means 2401 refers to an element <mediaobject> of structure description data through use of an "id" and reads a file name on the basis of attribute "locator" of the element <mediaobject>. Data pertaining to a file corresponding to the file name are read from the database, and the thus-read data are output to the demultiplex means 2402 (S2 and S3). The start time and the end time of the selected segment of the file, which are described so as to follow the "id," are output to the video skimming means 2403 and the audio skimming means 2404 (S4). After all the data sets have been subjected to the foregoing processing, processing is terminated (S5). If some of the data sets still remain unprocessed, the

54

previously-described processing is repeated after completion
of the processing effected by the demultiplex means 2402, the
processing effected by the video skimming means 2403, and the
processing effected by the audio skimming means 2404 (S6 and

5   S7).

Processing pertaining to the demultiplex means 2402 will
now be described. FIG. 50 is a flowchart showing processing
effected by the demultiplex means 2402. The demultiplex means
2402 receives a system stream of MPEG-1 format, which corresponds

10   to media content, from the interface means 2401 and separates
the thus-received system stream of MPEG-1 format into a video
stream and an audio stream. The video stream is output to the
video skimming means 2403, and the audio stream is output to
the audio skimming means 2404 (S1 to S10). After completion

15   of output of the video and audio streams (S9 and S11), termination
of the processing performed by the demultiplex means 2402 is
reported to the interface means 2401 (S12). As indicated by
the flowchart shown in FIG. 50, with the exception of transmission
of processing termination acknowledgement, the processing

20   performed by the demultiplex means 2402 is identical with that
performed by the demultiplex means according to the first
embodiment.

Processing effected by the video skimming means 2403 will
now be described. FIG. 53 is a flowchart showing the processing

25   effected by the video skimming means 2403. As indicated by the

55

flowchart shown in FIG. 53, with the exception of sending of

processing termination acknowledgement to the interface means

2401 performed at the end of the processing (S15 and S17), the

processing performed by the video skimming means 2403 is

5   identical with that effected by the video skimming means

according to the first embodiment.

Processing performed by the audio skimming means 2404 will

now be described. FIG. 52 is a flowchart showing the processing

effected by the audio skimming means 2402. As indicated by the

10   flowchart shown in FIG. 52, with the exception of sending of

a processing termination acknowledgement to the interface means

2401 at the end of processing (S11 and S12), the processing

performed by the audio skimming means is identical with that

performed by the audio skimming means described in connection

15   with the first embodiment.

In the formation step 1803, the video and audio streams

output in the extraction step 1802 are subjected to time-division

multiplexing by means of a multiplex method for MPEG-1

standardized under International Standard ISO/IEC IS 11172-1.

20   In a case where media content is stored into a plurality of separate

files, each of the files is multiplexed in the extraction step

1802 in order to output a video stream and an audio stream.

In the delivery step 1804, the system stream of MPEG-1 format

multiplexed in the formation step 1803 is delivered over the

25   line. When a plurality of system streams of MPEG-1 format are

output in the formation step 1803, all the system streams are delivered in the sequence in which they are output.

In the present embodiment, in a case where media content is stored into a plurality of separate files, each of the files is processed in the extraction step 1802. In the formation step 1803, wherein all the relevant video and audio streams of the files of media content are connected together and the thus-connected streams are output, the same advantageous result as that yielded in the formation step 1803 is achieved even when the video and audio streams are multiplexed into a single system stream of MPEG-1 format. In this case, the time code must be changed by the video skimming means 2403 such that the counter C for counting the number of output frames is incremented by only the amount corresponding to the number of video streams. The counter C is initialized at only the beginning of a file (S3 and S18 shown in FIG. 51). The processing effected by the video skimming means 2403 at this time is provided in the flowchart shown in FIG. 53. Although in the fifth embodiment the context description data and the physical context data are described separately from one another, these data sets may be merged into a single data set by means of appending attributes "seq" and "locator" of the structure description data to the attribute of the element <mediaobject> of the context description data.

[Sixth Embodiment]

A sixth embodiment of the present invention will now be

57

described by reference to the accompanying drawings. In the present embodiment, a motion picture of MPEG-1 format is taken as media content. In this case, a media segment corresponds to a single scene cut. Further, a score corresponds to the objective degree of contextual importance of a scene of interest.

FIG. 54 is a block diagram showing a media processing method according to the sixth embodiment of the present invention. In FIG. 54, reference numeral 3101 designates a selection step; 3102 designates an extraction step; 3103 designates a formation step; 3104 designates a delivery step; and 3105 designates a database. In the selection step 3101, a scene of media content is selected from context description data, and there are output data pertaining to the start time and the end time of the thus-selected scene, as well as data representing a file where the data are stored. Thus, processing pertaining to the selection step 3101 is identical with that effected in the selection step in the fifth embodiment. In the extraction step 3102, there are received the data sets representing the start time and the end time of the scene and the data representing the file, which are output in the selection step 3101. Further, data pertaining to the segment defined by the start and end time output in the selection step 3101 are extracted from the file of media content, by reference to structure description data. Processing pertaining to the extraction step 3102 is identical with that effected in the extraction step in the fifth embodiment.

58

In the formation step 3103, a portion or the entirety of the stream output in the extraction step 3102 is multiplexed according to the traffic volume determined in the delivery step 3104, thereby constituting a system stream of MPEG-1 format.

5 In the delivery step 3104, the traffic volume of the line over which the system stream of MPEG-1 format is delivered is determined, and the determination result is transmitted for use in the formation step 3103. Further, in the delivery step 3104, the system stream of MPEG-1 format prepared in the formation

10 step 3103 is delivered over the line. Reference numeral 3105 designates a database where media content, structure description data thereof, and context description data are stored.

FIG. 55 is a block diagram showing processing performed during the formation step 3103 and the delivery step 3104

15 according to the sixth embodiment. In FIG. 55, the formation step 3103 is embodied by stream selection means 3201 and multiplex means 3202. The delivery step 3104 is embodied by traffic volume determination means 3203 and delivery means 3204. The stream selection means 3201 receives the video and audio streams output

20 in the extraction step 3102 and the traffic volume output from the traffic volume determination means 3203. If the traffic volume of the line is sufficiently low to allow transmission of all data sets, all the system streams are output to the multiplex means 3202. If a long time is required for transmitting all

25 the data sets due to the line being busy or high traffic volume,

59

only portions of the plurality of audio and video streams are selected and output to the multiplex means 3202. In this case, selection may be implemented in several ways; namely, selection of only the basic layer of the video stream, selection of only monophonic sound of the audio stream, selection of only the left stereo signal of the same, selection of only the right stereo signal of the same, or like selection of a combination thereof. Here, if only a single video stream and a single audio stream exist, the streams are output regardless of the traffic volume. The multiplex means 3202 subjects the video and audio streams output from the stream selection means 3201 to time-division multiplexing, by means of the multiplex method for the MPEG-1 format standardized under International Standard ISO/IEC IS 11172-1. The traffic volume determination means 3203 examines the current state and traffic volume of the line over which streams are transmitted and outputs the results of examination to the stream selection means 3201. The delivery means 3204 delivers over the line the system stream of MPEG-1 format multiplexed by the multiplex means 3202.

In the present embodiment, in a case where a single video stream exists, the stream selection means 3201 outputs the video stream regardless of traffic volume. However, if transmission, over the line, of all the data sets pertaining to the video stream requires a large amount of time, only a representative image of the video stream may be selected and transmitted. At the

time of selection of a representative image, a time code of the representative image is described in the context description data. Alternatively, only a single frame, which is called I picture and can be decoded independently, may be selected from

5   among a plurality of frames.

[Seventh Embodiment]

A seventh embodiment of the present invention will now be described by reference to the accompanying drawings. In the present embodiment, a motion picture of a system stream of MPEG-1

10   format is taken as media content. In this case, a media segment corresponds to a single scene cut. Further, in the present embodiment, a score corresponds to the objective degree of contextual importance of a scene of interest from the viewpoint of a keyword related to a character or event selected by the

15   user.

FIG. 56 is a block diagram showing a processing method according to the seventh embodiment of the present invention. In FIG. 56, reference numeral 3301 designates a selection step; and 3302 designates an extraction step. In the selection step

20   3301, a scene of media content is selected from context description data by means of a keyword and a score thereof appended to the context description data. Data pertaining to the start time and the end time of the thus-selected scene are output. In the extraction step 3302, data pertaining to the segment

25   defined by the start time and end time output in the selection

61

step 3301 are extracted.

FIG. 57 shows the configuration of the context description data according to the seventh embodiment. In the present embodiment, the context is described according to a tree structure. Elements within the tree structure are arranged in chronological sequence from left to right. In FIG. 57, the root of the tree designated <contents> represents a single portion of content, and the title of the content is assigned to the root as an attribute.

Children of <contents> are designated by <section>. A keyword representing the contents or characters of a scene and priority representing the degree of importance of the keyword are appended to the element <section> as an attribute in the form of a pair of keyword and priority. The priority assumes an integral value ranging from 1 to 5, where 1 designates the least degree of importance and 5 designates the greatest degree of importance. The pair (a keyword and priority) is set so that it can be used as a key for retrieving a particular scene, or characters, as desired by the user. For this reason, a plurality of pairs (each pair including a keyword and priority) may be appended to a single element <section>. For example, in a case where characters are described, pairs are appended to a single element <section>, in a number equal to the number of characters appearing in a scene of interest. The value of the priority appended to the scene is set so as to become greater when a large

number of characters appear in a scene of interest.

Children of <section> are designated by <section> or . Here, an element <section> per se can be taken as a child of another child <section>. However, a single element <section> cannot have a mixture of children <section> and children .

An element represents a single scene cut. A pair (a keyword and priority) similar to that appended to the element <section> and time information about a scene of interest; namely, "start" representing the start time and "end" representing the end time, are appended to as attributes. Scenes may be cut through use of commercially-available software or software available over a network. Alternatively, scenes may be cut manually. Attribute "from" representing the start time of a scene can specify the start frame of a scene of interest. Although in the present embodiment time information is expressed in terms of the start time and the end time of a scene cut, a similar result is realized when time information is expressed in terms of the start time of the scene of interest and a duration of the scene of interest. In this case, the end time of the scene of interest is obtained by addition of the duration to the start time.

In the case of a story such as a movie, chapters, sections, and paragraphs can be described on the basis of the context description data, through use of elements <section>. In another

63

example, when a baseball game is described, elements <section> of the highest hierarchical level may be used for describing innings, and their children <section> may be used for describing half innings. Further, second-generation children <section>

5   of the elements <section> are used for describing at-bats of respective batters. Third-generation children<section>of the elements <section> are also used for describing each pitch, a time period between pitches, and batting results.

       The context description data having such a configuration

10   may be expressed in a computer through use of, e.g., Extensible Markup Language (XML). XML is a data description language whose standardization is pursued by the World Wide Web Consortium. Recommendations Ver. 1.0 was submitted on February 10, 1998. Specifications of XML Ver. 1.0 can be acquired from

15   http://www.w3.org/TR/1998/REC-xml-19980210. FIGS. 58 to 66 show one example of Document Type Definition (DTD) used for describing the context description data of the present embodiment through use of XML, and one example of context description data described through use of DTD. FIGS. 67 through 80 show one

20   example of context description data prepared by addition of representative data (dominant-data) of a media segment, such as a representative image (i.e., video data) and a keyword (audio data), to the context description data shown in FIGS. 58 through 66, and a DTD used for describing the context description data

25   through use of XML.

                                    64

Processing relating to the selection step S3301 will now be described. In the present embodiment, processing pertaining to the selection step S3301 is effected by focusing on an element and an element <section> having children . FIG. 81 is a flowchart showing processing pertaining to the selection step 3301 according to the seventh embodiment. In the selection step 3301, the keyword, which serves as a key for selecting a scene, and the threshold value of priority thereof are entered, thereby selecting an element <section> which has a keyword identical with the entered key and whose priority exceeds the threshold value from among elements <section> having elements of context description data as children (S2 and S3). Subsequently, only a child which has a keyword identical with the key and whose priority exceeds the threshold value is selected from among the children of the thus-selected element <section> (S5 and S6). The start time and end time of the selected scene are determined from attributes "start" and "end" of the child selected through the foregoing processing, and the start time and end time are output (S7, S8, S9, S10, S11, S1, and S4).

Although in the present embodiment selection is effected by focusing on an element and an element <section> having children , selection may be effected by focusing on another parent-and-child relationship; e.g., an element <section> and its child <section> within a certain hierarchical

65

stratum. Further, the parent-and-child relationship is not limited solely to a two-layer hierarchical stratum. The number of hierarchical levels of the hierarchical stratum may be increased to more than two, and leaves of the tree structure;

5 i.e., descendant , may be subjected to the same processing. Furthermore, the retrieval key may be set as a pair including a plurality of keywords and conditions defining the relationship between the keywords. Conditions defining the relationship between the keywords comprise combinations, such

10 as "either," "both," or "either or both." The threshold value for selection may be specified, and in the case of a plurality of keywords processing may be performed for each keyword. The keyword serving as a retrieval key may be entered by the user or automatically set by the system on the basis of a user profile.

15 Processing relating to the extraction step 3302 is identical with that effected in the extraction step described in connection with the first embodiment.

As shown in FIG. 82, the present embodiment yields an advantage of the ability to play back only scenes of media content

20 of interest as desired by an audience, by means of inputting the video stream output from the extraction step 3302 into video playback means and the audio stream output from the same into audio playback means, and playing back the audio and video streams, which are mutually synchronized. Further, there can be prepared

25 a system stream of MPEG-1 format relating to a collection of

scenes of media content of interest as desired by the audience, by means of multiplexing the video stream and the audio stream.

[Eighth Embodiment]

An eighth embodiment of the present invention will now be

5   described. The eighth embodiment differs from the seventh embodiment only in terms of the processing relating to the selection step.

Processing relating to the selection step S3301 will now be described. In the present embodiment, processing pertaining

10   to the election step S3301 is effected by focusing on only the element . FIG. 83 is a flowchart showing processing pertaining to the selection step 3301 according to the seventh embodiment. As shown in FIG. 83, in the selection step 3301, the keyword, which serves as a key for selecting a scene, and

15   the threshold value of priority thereof are entered. A child , which has a keyword identical with the key and whose priority exceeds the threshold value, is selected from among the elements of context description data (S1 to S6).

Although in the eighth embodiment selection is effected

20   by focusing on only the element , selection may also be effected by focusing on only an element <section> of a certain hierarchical level. Furthermore, the retrieval key may be set as a pair including a plurality of keywords and conditions defining the relationship between the keywords. Conditions

25   defining the relationship between the keywords comprise

combinations, such as "either," "both," or "either or both." The threshold value for selection may be specified, and in the case of a plurality of keywords processing may be performed for each keyword.

5   [Ninth Embodiment]

A ninth embodiment of the present invention will now be described. The ninth embodiment differs from the seventh embodiment only in terms of the processing relating to the selection step.

10   Processing relating to the selection step S3301 will now be described by reference to the accompanying drawings. As in the case of the processing described in connection with the seventh embodiment, in the selection step 3301 according to the ninth embodiment, selection is effected by focusing on only an

15   element and an element <section> having children . In the present embodiment, a threshold value is set with regard to the sum of duration periods of all scenes to be selected; more specifically, selection is effected such that the sum of the duration periods of the scenes that have been

20   selected so far is maximized but remains smaller than the threshold value. FIG. 84 is a flowchart showing processing relating to the selection step according to the ninth embodiment. In the selection step 3301, a single keyword, which serves as a retrieval key, is received. Subsequently, of the elements

25   <section> having children , all the elements <section>

68

having keywords identical with the retrieval key are extracted. A collection of the thus-selected elements <section> is taken as set Ω (S1 and S2). The elements <section> of the set Ω are sorted in descending order of priority (S3). Subsequently, the element <section> whose keyword or retrieval key has the highest priority value is selected from the thus-sorted elements of the set Ω (S5). The thus-selected element <section> is deleted from the set Ω (S6). In this case, if a plurality of elements <section> have the highest priority value, all the elements <section> are extracted. Of the children of the thus-selected elements <section>, only the children having the retrieval keys are selected, and the thus-selected children are added to another set Ω' (S7). The initial value of the set Ω' is "empty" (S2). The sum of duration periods of scenes pertaining to the set Ω' is obtained (S8), and the sum is compared with a threshold value (S9). If the sum of duration periods is equal to the threshold value, data pertaining to all the segments of the elements included in the set Ω' are output, and processing is terminated (S14). In contrast, if the sum of duration periods is lower than the threshold value, processing again returns to the selection from the set Ω (S5) of an element <section> whose retrieval key or keyword has the highest priority. The previously-described selection processing is repeated. If the set Ω is empty, data pertaining

69

to all the segments of the elements of the set Ω' are output, and processing is terminated (S4). If the sum of duration periods of the scenes relating to the set Ω' exceeds the threshold value, the following processing is performed. The

5   element whose retrieval key or keyword has the minimum priority is deleted from the set Ω' (S11). At this time, if a plurality of elements have the minimum priority, all the elements are deleted. The sum of duration periods of the elements of set Ω' is obtained (S12),

10   and the sum is compared with a threshold value (S13). If the sum of duration periods exceeds the threshold value, processing again returns to deletion of the elements from the set Ω' (S11). Such deletion processing is performed repeatedly. Here, if the set Ω' is empty, processing is terminated (S10).

15   In contrast, if the sum of duration periods is lower than the threshold value, data pertaining to all the segments of the elements of the set Ω' are output, and processing is terminated (S14).

Although in the present embodiment selection is effected

20   by focusing on an element and an element <section> having children , selection may be effected by focusing on another parent-and-child relationship; e.g., an element <section> and its children within another hierarchical level. Further, the parent-and-child relationship is not

70

limited solely to a two-layer hierarchical stratum; the number of hierarchical levels of the hierarchical stratum may be increased. For instance, in a case where elements are in the hierarchical layers ranging from an element <section> of the

5　highest hierarchical level to its child are subjected to processing, the element <section> of the highest hierarchical level is selected. Further, a successor <section> of the thus-selected element <section> is selected, and a second-generation child of the thus-selected element <section>

10　is further selected. Such a round of selection operations is repeated until the child is selected. The thus-selected elements are collected into a set Ω'.

In the present embodiment, elements are sorted in descending order of priority of the retrieval key or keyword. A threshold

15　value may be set with regard to the priority value, and elements may be selected in descending order of priority. The threshold value may be separately set with regard to the element <section>, as well as with regard to the element .

In the present embodiment, the retrieval key is specified

20　as a single keyword. However, the retrieval key may be set as a pair including a plurality of keywords and conditions defining the relationship between the keywords. Conditions defining the relationship between the keywords comprise combinations, such as "either," "both," or "either or both." In this case, there

25　is required a rule for determining the priority of keywords used

71

in selection or deletion of elements <section> and elements . One example of such a rule is as follows: If the condition is "either," the highest priority value of the priority values of corresponding keywords is set as "priority." Further, if the condition is "both," the minimum priority value of the priority value of corresponding keywords is set as "priority." Even when the condition is "either or both," the priority value can be determined in accordance with this rule. Further, in a case where a plurality of retrieval keys or keywords exist, a threshold value may be set with regard to the priority of the keywords as the retrieval keys, and elements whose priority value exceeds the threshold value may be processed.

[Tenth Embodiment]

A tenth embodiment of the present invention will now be described. The tenth embodiment differs from the seventh embodiment only in terms of the processing relating to the selection step.

Processing relating to the selection step S3301 will now be described by reference to the accompanying drawings. As in the case of the processing described in connection with the eighth embodiment, in the selection step 3301 according to the tenth embodiment, selection is effected by focusing on only an element . Further, as in the case of the ninth embodiment, in the present embodiment a threshold value is set with regard to the sum of duration periods of all scenes to be selected.

72

Specifically, an element is selected such that the sum of duration periods of scenes which have been selected so far is maximized but remains lower than the threshold value. FIG. 85 is a flowchart showing processing relating to the selection step according to the tenth embodiment.

In the selection step 3301, a single keyword, which serves as a retrieval key, is received. The set Ω' is initialized to "empty" (S2). Subsequently, of the elements , all the elements having keywords identical with the retrieval key are extracted (S1). A collection of the thus-selected elements is taken as set Ω. Subsequently, the elements whose keyword as the retrieval key has the highest priority value are sorted in descending order of priority (S3). From the thus-sorted elements of the set Ω, the element whose retrieval key as the keyword has the highest priority value is extracted (S5), and the thus-extracted element is deleted from the set Ω. In this case, if a plurality of elements have the highest priority value, all the elements are selected. If the set Ω is empty, data pertaining to all the segments of the elements of the set Ω' are output, and processing is terminated (S4). A sum, T1, of duration periods of the thus-extracted elements is computed (S6), and a sum, T2, of duration periods of scenes of the set Ω' is computed (S7). The sum of T1 and T2 is compared

73

with the threshold value (S8). If the sum of T1 and T2 exceeds
the threshold value, data pertaining to all the segments of the
elements included in the set Ω' are output, and
processing is terminated (S11). If the sum of T1 and T2 equals

5   the threshold value, all the extracted elements are
added to the elements of the set Ω' (S9 and S10), data pertaining
to all the segments of the elements included in the
set Ω' are output, and processing is terminated (S11). In
contrast, if the sum of T1 and T2 is lower than the threshold

10 value, all the extracted elements are added to the
elements of the set Ω', and processing then returns to selection
of elements from the set Ω (S10).

Although in the present embodiment selection is effected
by focusing on the elements , selection may be effected

15 by focusing on elements <section> in another hierarchical level.
In the present embodiment, elements are sorted in descending
order of priority of the keyword as the retrieval key. A
threshold value may be set with regard to the priority value,
and elements may be selected in descending order of priority,

20 given that the priority values of the elements are greater than
the threshold value.

Further, in the present embodiment, the retrieval key is
specified as a single keyword. However, the retrieval key may
be set as a pair including a plurality of keywords and conditions

74

defining the relationship between the keywords. Conditions

defining the relationship between the keywords comprise

combinations, such as "either," "both," or "either or both."

In this case, there is required a rule for determining the priority

5   of keywords used in selection or deletion of elements <section>

and . One example of such a rule is as follows: If

the condition is "either," the highest priority value of the

priority values of corresponding keywords is set as "priority."

Further, if the condition is "both," the minimum priority value

10   of the priority value of corresponding keywords is set as

"priority." Even when the condition is "either or both," the

priority value can be determined in accordance with this rule.

Further, in a case where a plurality of retrieval keys or keywords

exist, a threshold value may be set with regard to the priority

15   of the retrieval keys or keywords, and elements whose priority

value exceed the threshold value may be processed.

[Eleventh Embodiment]

An eleventh embodiment of the present invention will now

be described. The context description data of the present

20   embodiment differs from those of the seventh through tenth

embodiments, in terms of a viewpoint—which serves as a keyword

to be used for selecting a scene—and the description of degree

of importance of the viewpoint. As shown in FIG. 57, in the

seventh through tenth embodiments, the viewpoint and the degree

75

of importance based thereon are described by assigning a combination of a keyword and the degree of importance; i.e., (keyword, priority), to an element <section> or . In contrast, as shown in FIG. 133, according to the eleventh

5   embodiment, the viewpoint and the degree of importance thereof are described by assigning an attribute "povlist" to the root <contents> and assigning an attribute "povvalue" to an element <section> or .

As shown in FIG. 134, the attribute "povlist" corresponds

10   to a viewpoint expressed in the form of a vector. As shown in FIG. 135, the attribute "povvalue" corresponds to the degree of importance expressed in the form of a vector. Combination sets, each set comprising a viewpoint and the degree of importance thereof in a one-to-one relationship, are arranged in sequence

15   given, thus forming the attributes "povlist" and "povvalue." For instance, in illustrations shown in FIGS. 134 and 135, the degree of importance pertaining to viewpoint 1 assumes a value of 5, the degree of importance pertaining to viewpoint 2 assuming a value of 0; the degree of importance pertaining to viewpoint

20   3 assuming a value of 2; and the degree of importance pertaining to viewpoint "n" (where "n" designates a positive integer) assuming a value of 0. In the case of the seventh embodiment, the degree of importance pertaining to viewpoint 2 assuming a value of 0 means that viewpoint 2 is not assigned a keyword;

25   i.e., a combination (keyword, priority).

76

FIGS. 136 to 163 and FIGS. 164 to 196 show examples of

Document Type Definition (DTD) used for describing the context

description data of the present embodiment, through use of

Extensible Markup Language (XML) to be used in expressing the

5   context description data in a computer, and an example of context

description data described in DTD. Even in the present

embodiment, those processing operations which are the same as

those described in connection with the seventh through tenth

embodiments are effected through use of the context description

10   data.

In the present embodiment, the attribute "povlist" is

assigned to the root <contents>, and the attribute "povvalue"

is appended to an element <section> or . As shown in

FIG. 197, the attribute "povlist" may also be appended to an

15   element <section> or . In the case of an element

<section> or assigned the attribute "povlist," the

attribute "povvalue" corresponds to the attribute "povlist"

assigned to the element <section> or . In the case

of the element <section> or which is not assigned the

20   attribute "povlist," the attribute "povvalue" corresponds to

the attribute "povlist" assigned to the root <contents> or the

attribute "povlist" of the closest element <section> assigned

the attribute "povlist" from among the ancestors of an element

<section> or which is not assigned the attribute

25   "povlist."

77

FIGS. 198 to 252 show an example of DTD which corresponds to that shown in FIG. 197 and is used for describing the context description data of the present embodiment through use of XML to be used in expressing the context description data in a computer, and an example of context description data described in DTD. In these illustrated examples, the attribute "povvalue" assigned to an element <section) or corresponds to the attribute "povlist" assigned to the root <contents>.

[Twelfth Embodiment]

A twelfth embodiment of the present invention will now be described by reference to the accompanying drawings. In the present embodiment, a motion picture of a system stream of MPEG-1 format is taken as media content. In this case, a media segment corresponds to a single scene cut.

FIG. 86 is a block diagram showing a media processing method according to the twelfth embodiment of the present invention. In FIG. 86, reference numeral 4101 designates a selection step; 4102 designates an extraction step; 4103 designates a formation step; 4104 designates a delivery step; and 4105 designates a database. In the selection step 4101, a scene of media content is selected from context description data, and there are output data pertaining to the start time and the end time of the thus-selected scene, as well as data representing a file where the data are stored. In the extraction step 4102, there are received the data sets representing the start time and the end

78

time of the scene and the data sets representing the file output in the selection step 4101. By reference to the structure description data, data pertaining to the segment defined by the start and end time received in the selection step 4101 are extracted from the file of media content. In the formation step 4103, the data output in the extraction step 4102 are multiplexed, thus configuring a system stream of MPEG-1 format. In the delivery step 4104, the system stream of MPEG-1 format prepared in the formation step 4103 is delivered over a line. Reference numeral 4105 designates a database where media content, structure description data thereof, and context description data are stored.

The configuration of structure description data employed in the twelfth embodiment is identical with that described in connection with the fifth embodiment. More specifically, the structure description data having a configuration shown in FIG. 37 are used.

FIG. 87 shows the configuration of the context description data according to the twelfth embodiment. The context description data of the present embodiment corresponds to the context description data of the seventh embodiment appended a link to the element <mediaobject> of the structure description data. More specifically, the root <contents> of the context description data has a child <mediaobject>, and the element <mediaobject> has a child <section>. Elements <section> and

are identical with those used in the seventh embodiment. The element <mediaobject> of the context description data is appended an attribute "id." The element <mediaobject> of the structure description data is associated with the element

5 <mediaobject> of the context description data, by means of the attribute "id." Scenes of the media content described by means of decendants of the element <mediaobject> of the context description data are stored in a file designated by the element <mediaobject> of the structure description data having an

10 attribute id of the same value. Further, time information "start" and "end" assigned to an element "segment" set the time which has elapsed from the head of each file. Specifically, in a case where a single piece of media content comprises a plurality of files, the time at the head of each file corresponds

15 to 0, and the start time of each scene is represented by the time which has elapsed from the head of the file to a scene of interest.

The structure description data and the context description data may be expressed in a computer through use of, e.g.,

20 Extensible Markup Language (XML). FIG. 39 used in connection with the fifth embodiment shows one example of the structure description data. Further, FIGS. 88 to 96 show one example of Document Type Definition (DTD) used for describing the context description data shown in FIG. 87 through use of XML, and one

25 example of context description data described through use of

80

the DTD.

Processing relating to the selection step 4101 will now
be described. In the selection step 4101, any one of the methods
described in connection with the seventh through tenth

5 embodiments is adopted as a method of selecting a scene. The
"id" of the element <mediaobject> of corresponding structure
description data is eventually output simultaneously with output
of the start time and the end time of a selected scene. In a
case where the structure description data are described in the

10 form of an XML document through use of the DTD shown in FIG.
39 and where the context description data are described in the
form of an XML document through use of the DTD shown in FIGS.
88 and 96, one example of data output from the selection step
4101 is the same as that shown in FIG. 46 in connection with

15 the fifth embodiment.

Processing relating to the extraction step 4102 is identical
with the extraction step described in connection with the fifth
embodiment. The processing relating to the formation step 4103
is also identical with the formation step described in connection

20 with the fifth embodiment. Further, processing pertaining to
the delivery step 4104 is also identical with the delivery step
described in connection with the fifth embodiment.
[Thirteenth Embodiment]

A thirteenth embodiment of the present invention will now

25 be described by reference to the accompanying drawings. In the

present embodiment, a motion picture of a system stream of MPEG-1 format is taken as media content. In this case, a media segment corresponds to a single scene cut.

FIG. 97 is a block diagram showing a media processing method according to the thirteenth embodiment of the present invention. In FIG. 97, reference numeral 4401 designates a selection step; 4402 designates an extraction step; 4403 designates a formation step; 4404 designates a delivery step; and 4405 designates a database. In the selection step 4401, a scene of media content is selected from context description data, and there are output data pertaining to the start time and the end time of the thus-selected scene, as well as data representing a file where the data are stored. Processing relating to the selection step 4401 is identical with that relating to the selection step described in connection with the twelfth embodiment. In the extraction step 4402, there are received the data sets representing the start time and the end time of the scene and the data sets representing the file output in the selection step 4401. By reference to the structure description data, data pertaining to the segment defined by the start and end time received in the selection step are extracted from the file of media content. Processing relating to the extraction step 4402 is identical with that relating to the extraction step described in connection with the twelfth embodiment. In the formation step 4403, a portion or the entirety of the system stream output

82

in the extraction step 4402 is multiplexed in accordance with the traffic volume of the line determined in the delivery step 4404, thus configuring the system stream of MPEG-1 format. Processing relating to the formation step 4403 is identical with

5 that relating to the extraction step described in connection with the sixth embodiment. In the delivery step 4404, the traffic volume of the line is determined, and the determination result is transmitted to the formation step 4403. Further, the system stream of MPEG-1 format prepared in the formation step 4403 is

10 delivered over the line. Processing relating to the formation step 4404 is identical with that relating to the formation step described in connection with the sixth embodiment. Reference numeral 4405 designates a database where media content, structure description data thereof, and context description data are

15 stored.

Although in the thirteenth embodiment the system stream of MPEG-1 is taken as media content, the same advantageous result as that yielded by the MPEG-1 system stream can be yielded even by use of another format, so long as the format permits obtaining

20 of a time code for each screen.

Embodiments, which will be provided below, describe abstracts of modes corresponding to the inventions claimed in appended claims. An expression "sound data" will be hereinafter used as data pertaining to sound comprising audible tones,

25 silence, speech, music, tranquility, external noise or like sound.

83

An expression "video data" will be hereinafter used as data which are audible and visible, such as a motion picture, a static image, or characters such as telops. An expression "score" will be hereinafter used as a score to be calculated from the contents

5   of sound data, such as audible tones, silence, speech, music, tranquillity, or external noise; a score to be assigned in accordance with presence or absence of telops in the video data; or a combination thereof. Further, a score other than those mentioned above may also be used.

10   [Fourteenth Embodiment]

A fourteenth embodiment of the present invention will now be described and relates to an invention described in claim 2. FIG. 98 is a block diagram showing processing pertaining to a data processing method of the present embodiment. In the drawing,

15   reference numeral 501 designates a selection step; and 503 designates an extraction step. In the selection step 501, at least one segment or scene of media content is selected on the basis of a score of context description data, and the thus-selected segment or scene is output. The selected segment

20   corresponds to, for example, the start time and end time of a selected segment. In the extraction step 503, only the data pertaining to a segment of media content (hereinafter referred to as a "media segment") partitioned by the segment selected in the selection step S501; namely, the data pertaining to the

25   selected segment, are extracted.

Particularly, in the invention described in claim 5, a score corresponds to the objective degree of contextual importance of a scene of interest from the viewpoint of a keyword related to a character or event selected by the user.

[Fifteenth Embodiment]

A fifteenth embodiment of the present invention will now be described and relates to an invention described in claim 3. FIG. 99 is a block diagram showing processing pertaining to a data processing method of the present embodiment. In the drawing, reference numeral 501 designates a selection step; and 505 designates a playback step. In the playback step 505, only the data pertaining to the segment partitioned by a selected segment output in the selection step 501 are played back. Processing pertaining to the selection step 501 is the same as that described in connection with the first through thirteenth embodiments, and hence repetition of its explanation is omitted here for brevity.

[Sixteenth Embodiment]

A sixteenth embodiment of the present invention will now be described and relates to an invention described in claim 12. FIG. 100 is a block diagram showing processing pertaining to a data processing method of the sixteenth embodiment. In the drawing, reference numeral 507 designates a video the selection step; and 509 designates an audio selection step. Both the video the selection step 507 and the audio selection step 509

85

are included in the selection step 501 described in connection with the fourteenth and fifteenth embodiment.

In the video the selection step 507, a segment or scene of video data is selected by reference to context description data pertaining to video data, and the thus-selected segment is output. In the audio the selection step 509, a segment of sound is selected by reference to context description data pertaining to sound data, and the thus-selected segment is output. Here, the selected segment corresponds to, for example, the start time and end time of the selected segment. In the extraction step 503 described in connection with the fourteenth embodiment, only data from the segment of video data selected in the video the selection step 507 are played back. In the playback step 505, only data from the segment of sound data selected in the audio selection step 509 are played back.

[Seventeenth Embodiment]

A seventeenth embodiment of the present invention will now be described and relates inventions described in claims 15, 16, 17, 18, 19, and 20. FIG. 101 is a block diagram showing processing relating to a data processing method of the present embodiment. In the drawing, reference numeral 511 designates a determination step; 513 designates a selection step; 503 designates an extraction step; and 505 designates a playback step.

(Example 1)

In an invention described in claim 15, media content

86

comprises a plurality of different media data sets within a single period of time. In the determination step 511, there are received structure description data which describes the configuration of data of the media content. In this step, data which are objects

5 of selection are determined on the basis of determination conditions, such as the capability of a receiving terminal, the traffic volume of a delivery line, and a user request. In the selection step 513, there are received the data which are determined to be an object of selection in the determination

10 step 511, the structure description data, and the context description data. Further, a media data set is selected from only the data which are determined to be the object of selection in the determination step 511. Since the extraction step 503 is identical with the extraction step described in connection

15 with the fourteenth embodiment and the playback step 505 is identical with the playback step described in connection with the fifteenth embodiment, repetition of their descriptions is omitted here. Media data comprise several data sets, such as video data, sound data, and text data. In the following

20 description of examples, media data comprise in particular at least one of video data and sound data.

In the present example, as shown in FIG. 102, within a single period of time of media content, different video data or sound data are assigned to channels, and the video data or sound data

25 are further assigned to a hierarchical set of layers. For

instance, a channel-1/layer-1 for transmitting a motion picture is assigned to video data having a standard resolution, and a channel-1/layer-2 is assigned to video data having a high resolution. A channel 1 for transmitting sound data is assigned

5 to stereophonic sound data, and a channel 2 is assigned to monophonic sound data. FIGS. 103 and 104 show one example of Document Type Definition (DTD) used for describing structure description data through use of XML, and one example of context description data described through use of DTD.

10 In a case where media content is formed of such channels and layers, processing pertaining to the determination step 511 of the present example will now be described by reference to FIGS. 105 to 108. As shown in FIG. 105, in step 101 a determination is made as to whether or not a user request exists. If in step

15 101 a user request is determined to exist, the user request is subjected to determination processing SR-A shown in FIG. 106.

In step 101, if no user request is determined to exist, processing proceeds to step S103, where another determination is made as to whether or not receivable data are video data only,

20 sound data only, or both video and sound data. If in step S103 receivable data are determined to be solely video data, determination processing SR-C pertaining to video data shown in FIG. 107 is executed. If receivable data are determined to solely sound data, determination processing SR-C pertaining to

25 sound data shown in FIG. 108 is executed. If both video and

audio data are receivable, processing proceeds to step S105.
In step S105, a determination is made as to the capability of
a receiving terminal for receiving video and audio data; for
example, video display capability, playback capability, and a

5   rate at which compressed data are decompressed. If the
capability of the receiving terminal is determined to be high,
processing proceeds to step S107. In contrast, if the capability
of the receiving terminal is determined to be low, processing
proceeds to step S109. In step S107, the traffic volume of a

10   line over which video data and sound data are to be transported
is determined. If the traffic volume of the line is determined
to be high, processing proceeds to step S109. If the traffic
volume of the line is determined to be low, processing proceeds
to step S111.

15   Processing pertaining to step S109 is executed when the
receiving terminal has low capability or the traffic volume of
the line is high. During the processing, the receiving terminal
receives video data having a standard resolution over the
channel-1/layer-1 and sound data over the channel 2. Processing

20   pertaining to step S111 is executed when the receiving terminal
has high capability or the traffic volume is low. During the
processing, the receiving terminal receives video data having
a high resolution over the channel-1/layer-2 and stereophonic
sound over the channel 1.

25   The determination processing SR-A pertaining to user

89

request shown in FIG. 106 will now be described. In the present example, the user request is assumed to select a video layer and a sound channel. In step S151, a determination is made as to whether or not the user requests video data. If in step S151 the user is determined to request video data, processing proceeds to step S153. If the user is determined not to request video data, processing proceeds to step S159. In step S153, a determination is made as to whether or not the user request for video data corresponds to selection of a layer 2. If YES is chosen in step S153, processing proceeds to step S155, where the layer 2 is selected as video data. If NO is chosen in step S153, processing proceeds to step S157, where a layer 1 is selected as video data. In step S159, a determination is made as to whether or not the user requests audio data. If in step S159 the user is determined to request audio data, processing proceeds to step S161. If the user is determined not to request audio data, processing is terminated. In step S161, a determination is made as to whether or not the user request for audio data corresponds to selection of a channel 1. If YES is chosen in step S161, processing proceeds to step S162, where the channel 1 is selected as audio data. If NO is chosen in step S161, processing proceeds to step S165, where the channel 2 is selected as audio data.

The determination processing SR-B pertaining to video data shown in FIG. 107 will now be described. In step S171, a determination is made as to the capability of a receiving terminal

90

for receiving video data. If the receiving terminal is
determined to have high capability, processing proceeds to step
S173. If the receiving terminal is determined to have low
capability, processing proceeds to step S175. In step S173,
5   the traffic volume of a line is determined. If the traffic volume
of the line is determined to be high, processing proceeds to
step S175. In contrast, if the traffic volume of the line is
determined to be low, processing proceeds to step S177.

Processing pertaining to step S175 is executed when the
10   receiving terminal has low capability or the traffic volume of
the line is high. During the processing, the receiving terminal
receives only video data having a standard resolution over the
channel-1/layer-1. Processing pertaining to step S177 is
executed when the receiving terminal has low capability or the
15   traffic volume of the line is low. During the processing, the
receiving terminal receives only video data having a high
resolution over the channel-1/layer-2.

The determination processing SR-C pertaining to sound data
shown in FIG. 108 will now be described. In step S181, a
20   determination is made as to the capability of a receiving terminal
for receiving audio data. If the receiving terminal is
determined to have high capability, processing proceeds to step
S183. If the receiving terminal is determined to have low
capability, processing proceeds to step S185. In step S183,
25   the traffic volume of a line is determined. If the traffic volume

91

of the line is determined to be high, processing proceeds to step S185. In contrast, if the traffic volume of the line is determined to be low, processing proceeds to step S187.

Processing pertaining to step S185 is executed when the receiving terminal has low capability or the traffic volume of the line is high. During the processing, the receiving terminal receives only monophonic audio data over the channel 2. Processing pertaining to step S187 is executed when the receiving terminal has low capability or the traffic volume of the line is low. During the processing, the receiving terminal receives only stereophonic sound data over the channel 1.

(Example 2)

An invention described in claim 16 differs from the invention described in example 1 (the invention described in claim 15) in only processing pertaining to the determination step S511. In the determination step S511, there are received structure description data which describe the configuration of data of the media content. In this step, on the basis of determination conditions, such as the capability of a receiving terminal, the traffic volume of a delivery line, and a user request, a determination is made as to whether only video data, only sound data, or both video and sound data are to be selected. Since the selection step 513, the extraction step 503, and the playback step 505 are identical with those described previously, repetition of their explanations is omitted here.

Processing pertaining to the determination step 511 of the present example will now be described by reference to FIGS. 109 and 110. As shown in FIG. 109, in step S201 a determination is made as to whether or not a user request exists. If in step S201 a user request is determined to exist, processing proceeds to step S203. If no user request is determined to exist, processing proceeds to step S205. In step S203, a determination is made as to whether or not the user requests solely video data. If YES is chosen in step S203, processing proceeds to step S253, where only video data are determined to be an object of selection. If NO is chosen in step S203, processing proceeds to step S207. In step S207, a determination is made as to whether or not the user requests only sound data. If YES is chosen in step S207, processing proceeds to step S255, where only sound data are determined to be an object of selection. If NO is chosen in step S207, processing proceeds to step S251, where both video and audio data are determined to be objects of selection.

In step S205, to which processing proceeds when no user request exists, a determination is made as to whether only video data, only sound data, or both video and sound data are receivable. If in step S205 only video data are determined to be receivable, processing proceeds to step S253, where only video data are determined to be an object of selection. If in step S205 only sound data are determined to be receivable, processing proceeds to step S255, where only sound data are determined to be an object

93

of selection. If in step S205 both video and sound data are determined to be receivable, processing proceeds to step S209.

In step S209, the traffic volume of the line is determined. If the traffic volume of the line is low, processing proceeds to step S251, where both video and sound data are determined to be objects of selection. If the traffic volume of the line is high, processing proceeds to step S211. In step S211, a determination is made as to whether or not data to be transported over the line include sound data. If YES is chosen in step S211, processing proceeds to step S255, where sound data are determined to be an object of selection. If NO is chosen in step S211, processing proceeds to step S253, where video data are determined to be an object of selection.

(Example 3)

In an invention according to claim 17, media content comprises a plurality of different video and/or sound data sets at a single period of time. In addition to a determination as to whether only video data, only sound data, or both video and sound data are to be selected, which is made in the determination step 511 of the second example (according to the invention defined in claim 16), in the determination step S511 of the third example a determination is made as to which one of video data sets/audio data sets is to be selected as an object of selection, on the basis of determination conditions, such as the capability of a receiving terminal, the traffic volume of a delivery line,

94

and a user request.   Since the selection step 513, the extraction step 503, and the playback step 505 are identical with those described previously, repetition of their explanations is omitted here.

5   As in the case of example 1, within a single period of time of media content, different video data or sound data are assigned to channels or layers. For instance, a channel-1/layer-1 for transmitting a motion picture is assigned to video data having a standard resolution, and a channel-1/layer-2 is assigned to

10   video data having a high resolution. A channel 1 for transmitting sound data is assigned to stereophonic sound data, and a channel 2 is assigned to monophonic sound data. FIGS. 103 and 104 show one example of Document Type Definition (DTD) used for describing structure description data through use of XML, and one example

15   of context description data described through use of DTD.

Processing pertaining to the determination step 511 of the third example will now be described by reference to FIGS. 111 to 113. As shown in FIG. 111, in the present example, as in the case of the determination made in the example 2, data which

20   are an object of selection are determined (object-of-selection determination SR-D). In step S301, the data determined through the object-of-selection determination processing SR-D are determined. In step S301, when only video data are determined to be an object of selection, processing pertaining to

25   determination processing SR-E relating to video data shown in

FIG. 112 is executed. In step S301, when only audio data are determined to be an object of selection, processing pertaining to determination processing SR-F relating to audio data shown in FIG. 113 is executed. In step S301, when both video and audio data are determined to be an object of selection, processing proceeds to step S303, where the capability of a receiving terminal for receiving video and audio data is determined. If the receiving terminal is determined to have high capability, processing proceeds to step S305. If the receiving terminal is determined to have low capability, processing proceeds to step S307, where the capability of a line, such as a transmission rate, is determined. If the line is determined to have high capability, processing proceeds to step S309. In contrast, if the line is determined to have low capability, processing proceeds to step S307. If the line is determined to have a high traffic volume, processing proceeds to step S307. If the line is determined to have a low traffic volume, processing proceeds to step S311.

Processing relating to step S307 is executed when the receiving terminal has low capability, the line has low capability, or the line has a high traffic volume. During the processing, the receiving terminal receives video data having a standard resolution over the channel-1/layer-1 and monophonic sound data over the channel 2. In contrast, processing relating to step S311 is executed when the receiving terminal high

96

capability, the line has high capability, or the line has a low

traffic volume. During the processing, the receiving terminal

receives video data having a high resolution over the

channel-1/layer-2 and stereophonic sound data over the channel

5   1.

The determination processing SR-E pertaining to video data

shown in FIG. 112 will now be described. In step S351, a

determination is made as to the capability of a receiving terminal

for receiving video data. If the receiving terminal is

10   determined to have high capability, processing proceeds to step

S353. If the receiving terminal is determined to have low

capability, processing proceeds to step S355. In step S353,

the capability of the line is determined. If the capability

of the line is determined to be high, processing proceeds to

15   step S357. In contrast, if the capability of the line is

determined to be low, processing proceeds to step S355. In step

S357, the traffic volume of the line is determined. If the

traffic volume of the line is determined to be high, processing

proceeds to step S355. In contrast, if the traffic volume of

20   the line is determined to be low, processing proceeds to step

S359.

Processing relating to step S355 is executed when the

receiving terminal has low capability, the line has low

capability, or the line has a high traffic volume. During the

25   processing, the receiving terminal receives only video data

97

having a standard resolution over the channel-1/layer-1. In contrast, processing relating to step S359 is executed when the receiving terminal high capability, the line has high capability, or the line has a low traffic volume. During the processing,

5 the receiving terminal receives only video data having a high resolution over the channel-1/layer-2.

The determination processing SR-F pertaining to audio data shown in FIG. 113 will now be described. In step S371, a determination is made as to the capability of a receiving terminal

10 for receiving audio data. If the receiving terminal is determined to have high capability, processing proceeds to step S373. If the receiving terminal is determined to have low capability, processing proceeds to step S375. In step S373, the capability of the line is determined. If the capability

15 of the line is determined to be high, processing proceeds to step S377. In contrast, if the capability of the line is determined to be low, processing proceeds to step S375. In step S377, the traffic volume of the line is determined. If the traffic volume of the line is determined to be high, processing

20 proceeds to step S375. In contrast, if the traffic volume of the line is determined to be low, processing proceeds to step S379.

Processing relating to step S375 is executed when the receiving terminal has low capability, the line has low

25 capability, or the line has a high traffic volume. During the

processing, the receiving terminal receives only monophonic audio data over the channel 2. In contrast, processing relating to step S379 is executed when the receiving terminal high capability, the line has high capability, or the line has a low traffic volume. During the processing, the receiving terminal receives only stereophonic audio data over the channel 1.

(Example 4)

In inventions described in claims 18 and 19, representative data pertaining to a corresponding media segment are added, as an attribute, to individual elements of context description data in the lowest hierarchical layer. Media content comprises a plurality of different media data sets at a single period of time. In the determination step S511, there are received structure description data which describe the configuration of data of the media content. In this step, a determination as to which one of the media data sets and/or representative data sets is taken as an object of selection is made on the basis of determination conditions, such as the capability of a receiving terminal, the traffic volume of a delivery line, the capability of the line, and a user request.

Since the selection step 513, the extraction step 503, and the playback step 505 are identical with those described previously, repetition of their explanations is omitted here. Media data comprise video data, sound data, or text data. In the present example, media data include at least one of video

99

data and sound data. In a case where representative data correspond to video data, the representative data include, for example, representative image data for each media segment or low-resolution video data. In a case where representative data correspond to audio data, the representative data include, for example, key-phrase data for each media segment.

As in the case of example 3, within a single period of time of media content, different video data or sound data are assigned to channels or layers. For instance, a channel-1/layer-1 for transmitting a motion picture is assigned to video data having a standard resolution, and a channel-1/layer-2 is assigned to video data having a high resolution. A channel 1 for transmitting sound data is assigned to stereophonic sound data, and a channel 2 is assigned to monophonic sound data.

Processing pertaining to the determination step 511 of the present example will now be described by reference to FIGS. 114 to 118. As shown in FIG. 114, in step S401 a determination is made as to whether or not a user request exists. If in step S401 a user request is determined to exist, determination processing SR-G relating to user request shown in FIG. 116 is executed.

If in step S401 no user request is determined to exist, processing proceeds to step S403, where a determination is made as to whether only video data, only sound data, or both video and sound data are receivable. If in step S403 only video data

100

are determined to be receivable, determination processing SR-H relating to video data shown in FIG. 117 is executed. In contrast, if only sound data are determined to be receivable, determination processing SR-I relating to audio data shown in FIG. 118 is executed. If both video and sound data are determined to be receivable, processing proceeds to step S405 shown in FIG. 115.

In step S405, the capability of the receiving terminal is determined. After execution of processing pertaining to step S405, there are performed, in the sequence given, processing pertaining to step S407 for determining the capability of the line and processing pertaining to step S409 for determining the traffic volume of the line. On the basis of the results of the processing operations performed in steps S405, S407, and S409, in the determination step S511 of the present example a determination is made as to channels or layers of video data or audio data to be received, or as to representative data to be received.

TABLE 1

CAPABILITY OF RECEIVING TERMINAL	CAPABILITY OF LINE	DOES A LINE HAVE A HIGH TRAFFIC VOLUME?	RECEIVED DATA
HIGH	HIGH	NO	VIDEO DATA: CHANNEL 1, LAYER 2 AUDIO DATA: CHANNEL 1 (S411)

101

HIGH	HIGH	YES	VIDEO DATA: CHANNEL 1, LAYER 1 AUDIO DATA: CHANNEL 1 (S413)
HIGH	LOW	NO	VIDEO DATA: CHANNEL 1, LAYER 1 AUDIO DATA: CHANNEL 2 (S413)
HIGH	LOW	YES	VIDEO DATA: CHANNEL 1, LAYER 1 AUDIO DATA: CHANNEL 2 (S415)
LOW	HIGH	NO	VIDEO DATA: CHANNEL 1, LAYER 1 AUDIO DATA: CHANNEL 2 (S415)
LOW	HIGH	YES	VIDEO DATA: REPRESENTATIVE DATA AUDIO DATA: CHANNEL 2 (S417)
LOW	LOW	NO	VIDEO DATA: REPRESENTATIVE DATA AUDIO DATA: CHANNEL 2 (S417)
LOW	LOW	YES	VIDEO DATA: REPRESENTATIVE DATA AUDIO DATA: REPRESENTATIVE DATA (S419)

Determination processing SR-G relating to a user request shown in FIG. 116 will now be described. In step S451, a

determination is made as to whether or not the user requests only video data. If YES is chosen in step S451, processing determination SR-H pertaining to video data is performed. If NO is chosen in step S451, processing proceeds to step S453.

5 In step S453, a determination is made as to whether or not the user requests only audio data. If YES is chosen in step S453, determination processing SR-I relating to audio data is performed. If NO is chosen in step S453, processing proceeds to step S405.

Determination processing SR-H relating to video data shown

10 in FIG. 117 will now be described. In step S461, a determination is made as to the capability of the receiving terminal. After execution of processing pertaining to step S461, there are performed, in the sequence given, processing pertaining to step S463 for determining the capability of the line and processing

15 pertaining to step S465 for determining the traffic volume of the line. After the processing operations pertaining to these steps S461, S463, and S465 have been completed, only video data are received over the channel-1/layer-2 during the determination processing SR-H pertaining to video data of the present example,

20 provided that the receiving terminal has high capability, the line has high capability, and the traffic volume of the line is low (step S471). In contrast, if the receiving terminal has low capability, the line has low capability, and the traffic volume of the line is high, only representative video data are

25 received (step S473). If none of the foregoing conditions are

satisfied, only video data are received over the channel-1/layer-1 (step S475).

Determination processing SR-I relating to audio data shown in FIG. 118 will now be described. In step S471, a determination is made as to the capability of the receiving terminal. After execution of processing pertaining to step S471, there are performed, in the sequence given, processing pertaining to step S473 for determining the capability of the line and processing pertaining to step S475 for determining the traffic volume of the line. After the processing operations pertaining to these steps S471, S473, and S475 have been completed, only audio data are received over the channel 1 during the determination processing SR-I pertaining to audio data of the present example, provided that the receiving terminal has high capability, the line has high capability, and the traffic volume of the line is low (step S491). In contrast, if the receiving terminal has low capability, the line has low capability, and the traffic volume of the line is high, only representative audio data are received (step S493). If none of the foregoing conditions are satisfied, only video data are received over the channel 2 (step S495).

(Fifth Example)

In an invention described in claim 20, on the basis of determination conditions, such as the capability of a receiving terminal, the capability of a delivery line, the traffic volume

of the line, and a user request, a determination is made as to which one of the entire data pertaining to a media segment, only representative data pertaining to a corresponding media segment, or either the entire data pertaining to a corresponding media segment or representative data is to be taken as an object of selection.

As in the case of example 4, representative data pertaining to a corresponding media segment are added, as an attribute, to individual elements of context description data in the lowest hierarchical layer. In a case where representative data correspond to video data, the representative data include, for example, representative image data for each media segment or low-resolution video data. In a case where representative data correspond to audio data, the representative data include, for example, key-phrase data for each media segment.

Processing pertaining to the determination step 511 of the present example will now be described by reference to FIGS. 119 to 121. As shown in FIG. 119, in step S501 a determination is made as to whether or not a user request exists. If in step S501 a user request is determined to exist, determination processing SR-J relating to user request shown in FIG. 121 is executed.

In step S501 no user request is determined to exist, processing proceeds to step S503, where a determination is made as to whether only representative data pertaining to a media

segment, only the entire data pertaining to the media segment, or both the representative data and the entire data pertaining to the media segment are receivable. If in step S503 only representative data are determined to be receivable, processing

5   proceeds to step S553 shown in FIG. 120, wherein only representative data are determined to be taken as an object of selection. If only entire data are determined to be receivable, processing proceeds to step S555, wherein only the entire data are determined to be taken as an object of selection. If both

10   the representative data and the entire data are determined to be receivable, processing proceeds to step S505.

In step S505, the capability of the line is determined. If the line is determined to have high capability, processing proceeds to step S507. In contrast, if the line is determined

15   to have low capability, processing proceeds to step S509. In each of steps S507 and S509, the traffic volume of the line is determined. In step S507, if the line is determined to have low traffic volume, processing proceeds to step S551, where both the entire data and the representative data are determined to

20   be taken as objects of selection. In step S509, the line is determined to have high traffic volume, processing proceeds to step S553, where representative data are taken as an object of selection. If in step S507 the line is determined to have high traffic volume and in step S509 the line is determined to have

25   high traffic volume, processing proceeds to step S555, where

106

the entire data are taken as an object of selection.

During determination processing SR-J relating to a user request, in step S601 a determination is made as to whether a user request corresponds to only representative data.. If YES is chosen in step S601, processing proceeds to step S553, where only representative data are taken as an object of selection. If NO is selected in step S601, processing proceeds to step S603, where a determination is made as to whether or not the user request corresponds to only the entire data. If YES is chosen in step S603, processing proceeds to step S555, where only the entire data are taken as an object of selection. If NO is chosen in step S603, processing proceeds to step S551, where both the entire data and the representative data pertaining to the media segment are taken as objects of selection.

[Eighteenth Embodiment]

An eighteenth embodiment of the present invention will now be described. The present embodiment is directed to an invention described in claim 22. FIG. 122 is a block diagram showing processing pertaining to a data processing method of the present embodiment. Particularly, the processing is related to the invention described in claim 2. In the drawing, reference numeral 501 designates a selection step; 503 designates an extraction step; and 515 designates a formation step. Since the selection step 501 and the extraction step 503 are identical with those described in connection with the fourteenth embodiment,

repetition of their explanations is omitted here.

In the formation step 515, a stream of media content is formed from the data pertaining to a selected segment extracted in the extraction step 503. Particularly, in the formation step 5 515 a stream is formed by multiplexing the data output in the extraction step 503.

[Nineteenth Embodiment]

A nineteenth embodiment of the present invention will now be described. The present embodiment relates to an invention 10 described in claim 23. FIG. 123 is a block diagram showing processing pertaining to a data processing method of the present embodiment. In the drawing, reference numeral 501 designates a selection step; 503 designates an extraction step; 515 designates a formation step; and 517 designates a delivery step. 15 Since the selection step 501 and the extraction step 503 are identical with those described in connection with the fourteenth embodiment, repetition of their explanations is omitted here. Further, the formation step 515 is identical with the formation step described in connection with the eighteenth embodiment, 20 and hence repetition of its explanation is omitted.

In the delivery step 517, the stream formed in the formation step 515 is delivered over a line. The delivery step 517 may include a step of determining the traffic volume of the line, and the formation step 515 may include a step of adjusting the 25 amount of data constituting the file, on the basis of the traffic

108

volume of the line determined in the delivery step 517.

[Twentieth Embodiment]

A twentieth embodiment of the present invention will now be described. The present embodiment relates to an invention described in claim 24. FIG. 124 is a block diagram showing processing pertaining to a data processing method of the present embodiment. In the drawing, reference numeral 501 designates a selection step; 503 designates an extraction step; 515 designates a formation step; 519 designates a recording step; and 521 designates a data recording medium. In recording step 519, the stream formed in the formation step 515 is recorded on the data recording medium 521. The data recording medium 521 is used for recording a media content, context description data pertaining to the media content, and structure description data pertaining to the media content. The data recording medium 521such as a hard disk, memory, or DVD-RAM and so on. Since the selection step 501 and the extraction step 503 are identical with those described in connection with the fourteenth embodiment, repetition of their explanations is omitted here. Further, the formation step 515 is identical with the formation step described in connection with the eighteenth embodiment, and hence repetition of its explanation is omitted.

[Twenty-first Embodiment]

A twenty-first embodiment of the present invention will now be described. The present embodiment relates to an invention

109

described in claim 25. FIG. 125 is a block diagram showing

processing pertaining to a data processing method of the present

embodiment. In the drawing, reference numeral 501 designates

a selection step; 503 designates an extraction step; 515

5   designates a formation step; 519 designates a recording step;

521 designates a data recording medium; and 523 designates a

data recording medium management step. In data recording medium

management step 523, the media content which has already been

stored and/or media content which is to be newly stored are

10   reorganized according to the available disk space of the data

recording medium 521. More specifically, in the data recording

medium management step 523, at least one of the following

processing operations is performed. When the available disk

space of the data recording medium 521 is small, a media content

15   to be newly stored is stored after having been subjected to edition.

Context description data and structure description data, both

pertaining to the media content which has already been stored,

are sent to the selection step 501. The media content and the

structure description data are sent to the extraction step 503.

20   The media content is reorganized, and the thus-reorganized

content is recorded on the data recording medium 521. Further,

the media content which has not been reorganized is deleted.

Since the selection step 501 and the extraction step 503

are identical with those described in connection with the

25   fourteenth embodiment, repetition of their explanations is

110

omitted here.   Further, the formation step 515 is identical with

the formation step described in connection with the eighteenth

embodiment, and hence repetition of its explanation is omitted.

Moreover, since recording step 519 and data recording medium

5   521 are identical with those described in connection with the

nineteenth embodiment, repetition of their explanations is

omitted here.

[Twenty-second Embodiment]

A twenty-second embodiment of the present invention will

10   now be described.   The present embodiment relates to an invention

described in claim 26.   FIG. 126 is a block diagram showing

processing pertaining to a data processing method of the present

embodiment.   In the drawing, reference numeral 501 designates

a selection step; 503 designates an extraction step; 515

15   designates a formation step; 519 designates a recording step;

521 designates a data recording medium; and 525 designates a

stored content management step.   In the stored content

management step 525, the media content which have already been

stored in the data recording medium 521 are reorganized according

20   to the period of storage of the media content.   More specifically,

the stored content management step 525 comprises steps of:

managing the media content stored in the data recording medium

521; sending context description data and physical content data,

which pertain to a media content which have been stored over

25   a predetermined period of time, to the selection step 501; sending

111

the media content and the structure description data to the extraction step 503; re-organizing the media content; recording the thus-reorganized media content onto the data recording medium 521; and deleting the media content which has not been

5   re-organized yet.

Since the selection step 501 and the extraction step 503 are identical with those described in connection with the fourteenth embodiment, repetition of their explanations is omitted here.   Further, the formation step 515 is identical with

10   the formation step described in connection with the eighteenth embodiment, and hence repetition of its explanation is omitted. Moreover, since recording step 519 and data recording medium 521 are identical with those described in connection with the nineteenth embodiment, repetition of their explanations is

15   omitted here.

In the previously-described fourteenth through twenty-second embodiments, the selection steps 501 and 513 can be embodied as selection means; the video the selection step 507 can be embodied as video selection means; the audio the

20   selection step 509 can be embodied as audio selection means; the determination step 511 can be embodied as determination means; the formation step 515 can be embodied as formation means; the delivery step 517 can be embodied as delivery means; the recording step 519 can be embodied as recording means; the data

25   recording medium management step 523 can be embodied as data

112

recording medium management means; and the stored content management step 525 can be embodied as stored content management means. There can be embodied a data processing device comprising a portion of these means or all of the means.

5   In the previous embodiments, the media content may include a data stream, such as text data, other than video and audio data. Further, individual steps of the previous embodiments may be embodied by storage, into a program storage medium, of a program for causing the computer to perform processing

10   pertaining to all or a portion of the steps in the form of software or through use of a hardware circuit specifically designed so as to exhibit the features of the steps.

In the program expression of the context description data capable of being subjected to the software processing using a

15   computer, when a view point added to a <section> or overlaps with a view point of another <section> or , as shown in Fig. 253, a view point may be added to one of the <sections> or <segments> and the another <section> or may be represented as being linked to the view point thus added.

20   Further, as shown in Fig. 254, a view point table formed by collecting all the view points expressed in the context description data may be configured as a child element of the <contents> which are the root in the data structure of the context description data and may be arranged in a manner that each of

25   the <sections> or <segments> is added with a set of a link to

the corresponding view point in the view point table and a score

for the view point thus linked.   According to such a configuration,

since it becomes easy to show a list of the view points having

been registered (hereinafter called as a view point list) to

5   a user in advance, the user can know the view points having been

registered in advance before requesting a preferred view point.

Thus, when a view point desired by a user exists in the view

point list, the user can request the view point through the

selecting procedure from the view point list.   In this respect,

10   the view point table may be configured not only as the child

element of the <contents> but also as a child element of the

<sections> or <segments>, or may be configured to be described

separately.

Further, as shown in Fig. 255, the view point table may

15   be arranged as a mixed type view point table wherein all the

view points are not expressed so as to be linked to the view

point table but only a part of the view points is linked to the

view point table.   In this case, all the view points expressed

in the context description data are not necessarily described

20   in the view point table and only the view points referred by

the links are registered in the view point table.

Further, when a view point table representing the list of

the view points existing in the context description data is

configured so as to be described separately as shown in Fig.

25   256, the view point list based on the view point table can be

114

shown to a user before the user requests a preferred view point. In this case, the user can know the view points existing in the context description data before requesting the preferred view point, and also can request the view point by the selecting

5   procedure from the view point list.

Further, as shown in Fig. 257, the context description data may be expressed in such a configuration that it is divided into a data structure portion describing the data structure and an attribute portion describing view points as attributes, links

10   to the data structure portion and scores thereof.   In Fig. 257, an upper portion (a) represents the data structure portion and a lower portion (b) represents the attribute portion. In this figure, although the data structure portion is described as the most simple configuration, the data structure portion may be

15   arranged in the same configuration as the aforesaid embodiment. Further, the attribute portion is arranged in a manner that, at every view point, links to the subject <sections> or <segments> and scores related to the links of the view point are unified as a set.

20   The data structure portion and the attribute portion may not be described in the same file but may be described in separate files. Further, in the context description data arranged to be divided into the data structure portion and the attribute portion, the selection step (selection means) selects the

25   or <section> based on the score at every subject view

115

point of the attribute portion. Further, as shown in Fig. 258,

the respective view points of the attribute portion and the

<sections> or <segments> of the data structure portion may be

coupled by bi-directional links. In this case, the selecting

5   process designating the view point by the selection step

(selection means) may also be performed by the method described

in the aforesaid embodiment.

Further, as shown in Fig. 259, the context description data

may be expressed in such a configuration that it is divided into

10   a data structure portion describing the data structure and an

attribute portion wherein at every view point the links to the

data structure portion are described in the order of the higher

score. However, according to such an expression, since there

is a possibility that fine comparison becomes impossible in the

15   case of comparing the scores among a plurality of the view points,

the rough order of "high", "middle" and "low" is used.

Furthermore, as shown in Fig. 260, the context description

data may be expressed in such a configuration that it is divided

into a data structure portion describing the data structure and

20   an attribute portion wherein at every view point the links to

the data structure portion are described in the order of the

higher scores and the links of the same score are described so

as to be arranged in two rows. In this case, the selecting process

designating the view point by the selection step (selection

25   means) may also be performed in the same manner as the processing

116

performed as to the context description data shown in Fig. 258.

Then, explanation will be made as to a context date converting method for converting the context description data of a tree structure into context description data (hereinafter called second context description data) which differs in the data structure from the context description data of a tree structure. In this respect, the context description data of a tree structure is arranged in a manner as shown in Fig. 57 that it is configured by <contents> as a root, <sections> as nodes and <segments> as leaves, each of the <sections> and <segments> is added as an attribute with at least one set (keyword, priority) of a "keyword" which is a keyword representing contents or a person etc. of a scene and a "priority" representing the degree of importance, and each of the <segments> is further added as time information of the scene with "start" representing start time and "end" representing end time or "duration" representing duration time.

In this specification, three kinds of context date converting methods will be explained. Hereinafter, the explanation will be made as to the data structures of the second context description data which are prepared by the respective context date converting methods and then examples of the respective context date converting methods will be explained.

(First embodiment of the context date converting method)

First, the second context description data prepared by the

117

first embodiment of the context date converting method is configured in a tree structure as shown in Figs. 261 and 262 by <contents> as a root, <keywords> as child elements of the <contents>, <levels> as child elements of each of the <keywords>

5   and <segments> as a child element of a <level>. In this respect, the element (node) of the <section> existing in the original context description data shown in Fig. 57 is not described in the second context description data. In the second context description data, the brother relationship of the tree structure

10   is also arranged in time sequence from the left side. Further, each of the <segments> is added with time information (start, end) which is originally added thereto.

     The <level> of the second context description data is determined based on the "priority" used in the context

15   description data shown in Fig. 57 and represents the degree of the importance. When the "priority" is expressed by an integer number, an integer number allotted to the "priority" is used as it is as the <level>. In contrast, when the "priority" is expressed by a decimal point, <level> is newly set in accordance

20   with a level of a value allotted to the "priority" so that the degree of the importance can be compared easily. For example, when three "priorities" of 0.2, 0.5 and 1.0 exist in the original context description data, <level 1> with the lowest degree of the importance is allotted to the "priority" of 0.2, <level 2>

25   with the middle degree of the importance is allotted to the

"priority" of 0.5, and <level 3> with the highest degree of the importance is allotted to the "priority" of 1.0.

In this manner, the levels of the degree of the importance are set, and the data structure of the second context description

5   data may be arranged in a nest form wherein the high-low relation of the <levels> is represented as a set relationship of the tree structure as it is as shown in Fig. 261, or may be arranged in a parallel form wherein the high-low relation of the <levels> is represented as the same layer as shown in Fig. 262. Further,

10   the data structure of the second context description data may be arranged in a manner as shown in Fig. 263 that the child elements of the <keyword> except for the <level> information are set as <segments> and the degrees of the importance may be determined in accordance with the coupling order to the <keyword>. In this

15   respect, since there is a possibility that fine comparison may not be performed at the time of comparing among a plurality of keywords, the comparison result may be represented roughly by "high", "middle" and "low", for example.

When a plurality of <segments> being continuously connected

20   exist among the <segments> belonging to the same <keyword> and the same <level>, these plurality of <segments> may be collected as one set. For example, when there are the and the being continuously connected, these <segments> may be collected as a . In this case, it is necessary

25   to prepare time information to be added to the thus

119

collected on the basis of the time information (start, end) added to the respective <segments>.

Hereinafter, the first embodiment of the context date converting method will be explained by using an example of the original context description data. As shown in Fig. 264, the original context description data as an example used in the following explanation includes as leaves the added with (keyword a, priority 2) and (keyword b, priority 1) as attributes, the added with (keyword a, priority 2) and (keyword c, priority 2), and the added with (keyword b, priority 4) and (keyword d, priority 3).

First, a set of "keywords" added to the <segments> of the original context description data is obtained. {keyword a, keyword b, keyword c, keyword d} are obtained from an example of the original context description data (hereinafter merely called as an example). Then, a set of <segments> added with the same keyword is obtained at every keyword. In the example, a set formed by the (priority 2) and the (priority 2) is obtained as to the keyword a, a set formed by the (priority 1) and the (priority 4) is obtained as to the keyword b, a set formed by the (priority 2) is obtained as to the keyword c, and a set formed by the (priority 3) is obtained as to the keyword d.

Then, the sets of the segments are divided into groups at

120

every priority.   For example, in the set of the segments of the keyword a, since each of the segments is added with the priority 2, these two segments are collected as a group of the priority 2.   As to the set of the segments of the keyword b, since the

5   is added with the priority 1 and the is added with the priority 4, these segments are divided into a group of the priority 1 (only the) and a group of the priority 4 (only the).   The segments of each of the keyword c and the keyword d are divided into groups in

10   the same manner.

Then, the respective priorities are converted into "levels" representing the degrees of the importance.   As described above, when the "priority" is expressed by an integer number, an integer number allotted to the "priority" is used as it is as the <level>.

15   Thus, in the aforesaid example, the added with the priority N (N=1, 2, 3, 4, 5) is the segment of the level N.

Further, when the nest form is employed as the second context description data, the respective segments being grouped are set as a set relationship in accordance with the levels of the

20   respective segments.   In contrast, when the parallel form is employed, the respective segments being grouped are set as a brother relationship and arranged in the order of the higher level or the lower level.

The data structure of the second context description data

25   thus prepared is shown in Fig. 265.   In this figure, since there

are the and the as the segments of the
<keyword a> and the <level 2>, these segments may be collected
as a , for example, when these segments are
continuously connected with no gap of time.

5   (Second embodiment of the context date converting method)

The second context description data prepared by the second
embodiment of the context date converting method is configured
in a tree structure as shown in Fig. 266 by <contents> as a root,
<keywords> as child elements of the <contents> and <segments>
10 as child elements of each of the <keywords>. In this respect,
each of the <segments> is added with a "priority" as an attribute.

In this second context description data, like the second
context description data prepared by the first embodiment of
the context date converting method, the <section> is not
15 described and each of the <segments> is added with time
information (start, end) in addition to the "priority". Since
the second embodiment of the context date converting method does
not convert the priority into the level, the context date
converting method of this embodiment does not describe the
20 <level> which is described in the second context description
data prepared by the first embodiment of the context date
converting method.

Hereinafter, the second embodiment of the context date
converting method will be explained by using an example of the
25 original context description data used above. First, like the

122

first embodiment of the context date converting method, a set of "keywords" added to the <segments> of the original context description data is obtained. Then, a set of <segments> added with the same keyword is obtained at every keyword. Then, each

5   of the <segments> is added with a priority which was added originally.

The data structure of the second context description data thus prepared is shown in Fig. 267. In this figure, since there are the and the as the segments of the

10 <keyword a>, these segments may be collected as a , for example, when these segments are continuously connected with no gap of time and the priorities added to these segments are same.

(Third embodiment of the context date converting method)

15 The second context description data prepared by the third embodiment of the context date converting method is configured in a tree structure as shown in Figs. 268 and 269 by <contents> as a root, <levels> as child elements of the <contents> and <segments> as child elements of each of the <levels>. In this

20 respect, each of the <segments> is added with a "keyword" as an attribute.

In this second context description data, like the second context description data prepared by the first embodiment of the context date converting method, the <section> is not

25 described and each of the <segments> is added with time

123

information (start, end) in addition to the "priority".

Hereinafter, the third embodiment of the context date converting method will be explained by using an example of the original context description data used above. First, a set of 5 <segments> added with the same priority is obtained at every priority based on the "priorities" added to the <segments> of the original context description data. In the example of the original context description data shown in Fig. 264, a set formed by the , the (keyword b) is obtained as to the priority 1, a set formed by the , the (keyword a) and a set 10 formed by the , the (keyword a, keyword c) are obtained as to the priority 2, a set formed by the , the (keyword d) is obtained as to the priority 3, and, a set formed by the , the (keyword b) is obtained as to the priority 4.

15 Then, the respective priorities are converted into "levels" representing the degrees of the importance. As described above, when the "priority" is expressed by an integer number, an integer number allotted to the "priority" is used as it is as the <level>. Thus, in this example, the added with the priority 20 N (N=1, 2, 3, 4, 5) is the segment of the level N.

Then, each of the segments is added with the "keyword" originally added to the corresponding priority. For example, the set of the segments of the level 1 has the and the is originally added with the keyword b, so that 25 this segment is added with the keyword b. The set of the segments

124

of the level 2 has the and the . In particular, since the added with the level 2 include the originally added with the keyword a and the originally added with the keyword c, another is newly prepared, and one of the two <segments 2> is added with the keyword a and the other of the <segments 2> is added with the keyword c. Similar processing is performed as to each of the of the level 2, the of the level 3 and the of the level 4.

The data structure of the second context description data thus prepared is shown in Fig. 270. In the case where the is added with a plurality of different keywords, the data structure may be arranged in a manner that the plurality of different keywords are added to the as shown in Fig. 269. Thus, in the second context description data shown in Fig. 270, the of the level 2 may be added with the priority a and the priority c. Further, in this figure, although the segments of the level 2 include the added with the priority a and the added with the priority a, these segments may be collected as a , for example, when these segments are continuously connected with no gap of time.

The second context description data prepared by the first to third embodiments of the context description data converting method explained above is used in the processes of S5 to S9 in

125

the flow chart of the selection step explained in the seventh

embodiment. Although the selection step using the original

context description data as shown in Fig. 57 has such flexibility

of capable of coping with any request, the selection step using

5   the second context description data is advantageous in that a

corresponding can be obtained quickly in response to

the request from a user.

Although in the aforesaid embodiments, the <section> or

having a value equal to or larger than the threshold

10   value is selected by using the threshold value of the degree

of the importance, the <section> or having a particular

vale of the degree of the importance may be selected.

Although in the previous embodiments context description

data and structure description data have been described

15   separately, they may be combined into a single data set, as shown

in FIGS. 127 to 132.

As has been described previously, according to the data

processing method, the recording medium, and the program of the

present invention, at least one segment is selected from a media

20   content on the basis of a score appended to context description

data by means of selection means (corresponding to the selection

step), through use of context description data of hierarchical

stratum. Particularly, only the data pertaining to a segment

selected by the selection means (corresponding to the selection

25   step) are extracted by means of the extraction means

126

(corresponding to the extraction step). Alternatively, only the data pertaining to the segment selected by the selection means (corresponding to the selection step) are played back, by means of the playback means (corresponding to the playback

5   step).

By means of the foregoing configuration, a more important scene can be freely selected from the media content, and the thus-selected important segment can be extracted or played back. Further, the context description data assume a hierarchical

10   stratum comprising the highest hierarchical layer, the lowest hierarchical layer, and other hierarchical layers. Scenes can be selected in arbitrary units, such as on a per-chapter basis or a per-section basis. There may be employed various selection formats, such as selection of a certain chapter and deletion

15   of unnecessary paragraphs from the chapter.

In the data processing method, the recording medium, and the program of the present invention, a score represents the degree of contextual importance of media content. So long as the score is set so as to select important scenes, a collection

20   of important scenes of a program, for example, can be readily prepared. Further, so long as the score is set so as to represent the importance of a scene of interest from the viewpoint of keyword, segments can be selected with a high degree of freedom by determination of a keyword. For example, so long as a keyword

25   is determined from a specific viewpoint, such as a character

127

or an event, only the scenes desired by the user can be selected.

In the data processing method, the recording medium, and the program of the present invention, in a case where media content comprises a plurality of different media data sets within a single

5 period of time, the determination means (corresponding to the determination step) determines which of the media data sets is to be taken as an object of selection, on the basis of determination conditions. The selection means (corresponding to the selection step) selects a media data set from only the data

10 determined by the determination means (corresponding to the determination step). Since the determination means (corresponding to the determination step) can determine media data pertaining to an optimum segment according to determination conditions, the selection means (corresponding to the selection

15 step) can select an appropriate amount of media data.

In the data processing method, the recording medium, and the program of the present invention, the determination means (corresponding to the determination step) determines whether only the video data, only the audio data, or both video and audio

20 data are to be taken as an object of selection, on the basis of the determination conditions. As a result, the time required by the selection means (corresponding to the selection step) for selecting a segment can be shortened.

In the data processing method, the recording medium, and

25 the program of the present invention, representative data are

128

appended to the context description data as an attribute, and the determination means can determine media data of an optimum segment or representative data, according to determination conditions.

5   In the data processing method, the recording medium, and the program of the present invention, the determination means (corresponding to the determination step) determines whether only the entire data pertaining to a corresponding media segment, only the representative data, or both the entire data and 10 representative data are to be taken as objects of selection, on the basis of the determination conditions. As a result, the determination means can shorten the time required by the selection means (corresponding to the selection step) for selecting a segment.